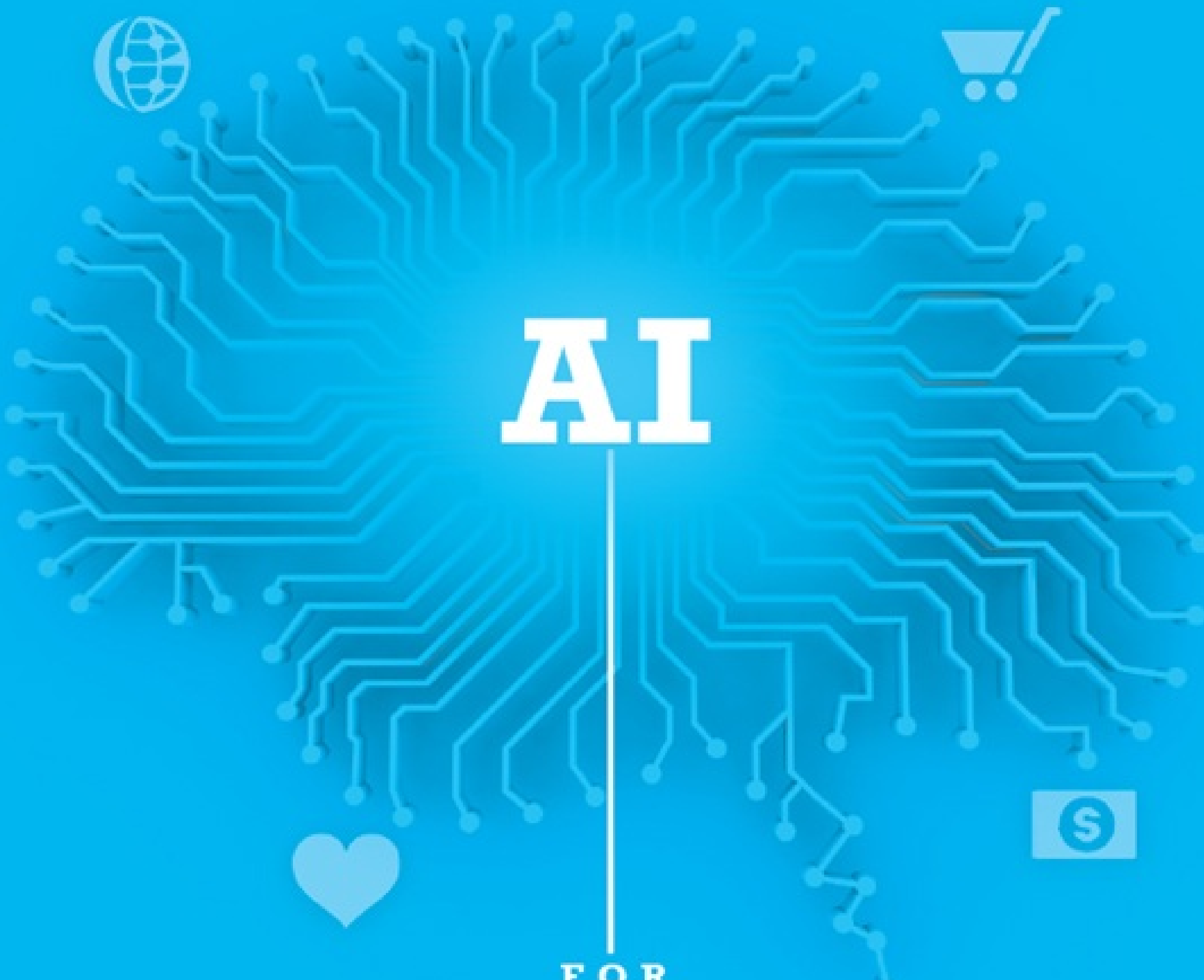
DR. A.K. PRADEEP | ANDREW APPEL | STAN STHANUNATHAN



AI FOR MARKETING AND PRODUCT INNOVATION

Powerful New Tools for Predicting Trends, Connecting with Customers, and Closing Sales

WILEY

DR. A.K. PRADEEP | ANDREW APPEL | STANS THANUNATHAN

AI

FOR

MARKETING

AND

PRODUCT INNOVATION

Powerful New Tools for Predicting Trends,
Connecting with Customers, and Closing Sales

WILEY

Cover images: © ALFRED PASIEKA/SCIENCE PHOTO LIBRARY/
Getty Images; © sodafish/Getty Images
Cover design: Wiley

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at http://booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

ISBN 9781119484066 (Hardcover)
ISBN 9781119484080 (ePDF)
ISBN 9781119484097 (ePub)

Dr. Pradeep *dedicates this book to his daughters, Alexis and Shane, and his son, Devin, who inspire him with their wit and humor, and to his fiancée, Mara, who inspires him with her love and boundless enthusiasm.*

Andrew Appel *dedicates this book to Anne, Sacha, Olivia, Henry, and Minna, who support his passion for working, and to his Mom, from whom he gets his boundless entrepreneurial energy.*

Stan Sthanunathan *dedicates this book to Aditya, Atishi, and Anusha for their love, and to Uma for her tolerance.*

CONTENTS

List of Tables

List of Illustrations

PREFACE

> *Everything that civilisation has to offer is a product of human intelligence; we cannot predict what we might achieve when this intelligence is magnified by the tools that AI may provide, but the eradication of war, disease, and poverty would be high on anyone's list. Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last.*
>
> – Stephen Hawking, *Independent*, May 1, 2014

Artificial Intelligence (AI) might mean the end of mankind?! Maybe we should head for the hills!

Or maybe the movie theater:

> Sony Pictures Animation announced *The Mitchells vs. The Machines*, an AI-as-evildoer animated family comedy.
>
> The story is: The Mitchells are a dysfunctional but loving family whose road trip is interrupted by a tech uprising that threatens mankind.
>
> All around the world, the electronic devices people love – from phones to self-driving cars to a sleek new line of personal robots – turn on humanity. With the help of two malfunctioning robots and the family's delightfully overweight pug, the Mitchells will have to get past their problems and work together to save each other and the world.

AI has entered the cultural bloodstream for sure when Hollywood is making animated family comedies about it!

So what is the reality: Doom and destruction – or delight in the darkness of the local cineplex?

The fact is, no one really knows for sure. That simple truth alone reflects the scale and the importance that AI is already playing in our lives – and if one thing IS certain, it is that AI and its associated technologies will only become more central as the future unfolds. Optimists and pessimists can and will continue to debate the prospects for that future, as they should. But on that overarching point, they would all agree.

This book focuses on ways in which AI can be harnessed for business applications: specifically, product innovation and marketing. Chapters touch upon a myriad of business-related subjects, from pricing and promotions to the future of market research and advertising agencies. The goal is to give you, the reader, an in-depth look at what AI is, what it can – and cannot – do, and provide ideas and insights on ways in which you might apply that knowledge to your own business and your own career.

This book is written to inspire the marketing professional or product innovator. To give you the reader enough of a grasp so you can be inspired to put the book down, and think of what you can do with it. The goal is not to create algorithmic mumbo-jumbo or a litany of case studies that do not seem peripherally applicable to your day job. The goal is to inspire you to "Think Different." You need not be a computer expert nor know your way around a line of code to extract value from the contents herein. You need only to want to gain a meaningful glimpse into the future of business, and understand how and why daily life around us will increasingly be conducted in close partnerships – seen and unseen – with the intelligent machines that are already among us.

For businesses large and small, global and local, the real question is: What are the practical implications that Artificial Intelligence and Machine Learning (ML) have for my company? How can they best be put to work to gain a competitive advantage in today's increasingly digitally driven economy? What do smart marketers want and need to know about the fascinating fields of AI and ML in order to understand and apply them to real-world business challenges?

This book is based on real-life examples of AI and ML at work. Techniques described in the book have become algorithms. The book describes the complementary disciplines of ML and AI so that readers can gain a better grasp of the new world we are already living in. This book outlines the resources, the skills, the best practices, the terminology, and the metrics required to harness the unparalleled and rapidly expanding power of these twin technologies.

But beyond serving as a marketer's primer on this most timely subject, this is also a book that seeks to encourage the creative community to embrace and employ AI and ML in ways that speak centrally to the human mind and spirit. There is a reason the word "elevate" is used in these pages.

The most effective sales methods, messages, and new product ideas are those that resonate most deeply and meaningfully with consumers at the non-conscious level. AI and ML can deliver tools that aid both marketers and creatives in discovering and developing those messages and product innovations.

The book is not a repeat of well-understood marketing techniques and terminology, but rather looks at the same areas through the lens of Machine Learning and Artificial Intelligence.

This is not a coffee-table book. This is a *desktop* book – designed to be an at-hand resource, a reliable guide, and a source of inspiration for successful product innovation and marketing in the exciting new Age of AI.

ACKNOWLEDGMENTS

From Dr. Pradeep

Profound and meaningful conversations with Professor George Lakoff at Berkeley, Professor Rajiv Lal at Harvard, and Professor Shankar Sastry at Berkeley.

Colleagues and friends at Unilever, Coca-Cola, Johnson & Johnson, Hersheys, Nike, Clorox, Red Bull, Chanel, Starbucks, Miller Coors, Home Depot, Rag & Bone, and ITC India, who contributed ideas, thoughts, and inspirations.

Daphne, Dev, Ying, David, Daniel, Mithun, Katy, Chris, Jared, Eric, Ellen, Luca, and Rachael in the United States, and Anirudh, Susmitha, Sathishkumar, Rajendran, Sugumar, Prem, and Jai in India, who make MachineVantage and its algorithms hum every day producing insights that fill the pages of this book.

Bob, Piyush, Ash, Jose Carlos, Krishnakumar, and Nishat at IRI for hours spent in debate and conversation over wine.

Tom Robbins for tinkering with thought itself.

Olivier, Ian, Andrew, Stan, Ben, and Russ who made all of this possible by investing in MachineVantage.

Wisdom and wit of my co-authors!

Sunrises on the island of Paros in Greece, great beer in London, wine in Napa, sunsets on Stinson Beach, and pizza in Chicago that inspired content in the book.

From Andrew Appel

Colleagues at IRI whose desire to disrupt industries makes cutting-edge innovation happen every day.

Ash Patel, whose technological genius is unmatched for teaching me everything about Big Data, AI, and Machine Learning.

My executive team, who every day works with our clients to help accelerate data-driven, machine-augmented decision-making in companies.

To the field teams of IRI who help consumers make better buying decisions.

Board members of New Mountain Capital for the vision and courage to hit the "Go Button" so many times as we race in the Age of the Algorithm.

From Stan Sthanunathan

Leadership team of CMI for embracing new ideas and concepts with an open mind.

Paul and Keith, for sponsorship and support of scholarship and professional growth.

Supplier ecosystem for partnership and sharing of knowledge.

Introduction

There is such a volume of information, speculation, hype, fear-mongering, and futuristic fantasies surrounding Artificial Intelligence. A simple description penned in 1968 by Stanley Kubrick and Arthur C. Clarke in *2001: A Space Odyssey*, with added comic relief from state-of-the-art systems, seems like a good place to start.

Dave Bowman: Hello, HAL. Do you read me, HAL?

Alexa: I cannot read you, would you like to buy a book?

Siri: Hello, HALLO – greetings.

HAL: Affirmative, Dave. I read you.

Dave Bowman: Open the pod bay doors, HAL.

Alexa: Coffee pods are on sale, $19.99/lb, shall I order one?

Siri: iPods once open will lose warranty.

HAL: I'm sorry, Dave. I'm afraid I can't do that.

Dave Bowman: What's the problem?

Alexa: There is no problem.

Siri: I am sorry, I don't understand that. Please try again.

HAL: I think you know what the problem is just as well as I do.

Dave Bowman: What are you talking about, HAL?

Alexa: What would you like to talk about?

Siri: Sorry for being obtuse.

HAL: This mission is too important for me to allow you to jeopardize it.

Dave Bowman: I don't know what you're talking about, HAL.

Alexa: I am sorry you don't know what you are talking about.

Siri: Don't worry about it.

HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.

Dave Bowman: Where the hell’d you get that idea, HAL?

Alexa: Hell hot sauce available to order. Shall I place the order?

Siri: I am afraid I don’t know the answer to that.

HAL: Dave, although you took very thorough precautions in the pod against my hearing you, I could see your lips move.

Dave Bowman: Alright, HAL. I’ll go in through the emergency airlock.

Alexa: In case of emergency call 911.

Siri: . . .

HAL: Without your space helmet, Dave, you’re going to find that rather difficult.

Dave Bowman: HAL, I won’t argue with you anymore. Open the doors.

Alexa: Door openers available on sale.

Siri: Sorry, I wasn’t able to find anything like that at this time.

HAL: Dave, this conversation can serve no purpose anymore. Goodbye.

Artificial Intelligence (AI) is a display of intelligence by a nonliving object, such as a machine, as opposed to Natural Intelligence, which is seen in living creatures, including humans. Artificial Intelligence itself is nothing new to the world of technology, having become officially recognized as an academic field way back in 1956.

What is new is the way people think about and actually experience Artificial Intelligence. We now recognize the virtually limitless practical applications of intelligent machines, because we interact with them on a daily basis, often in the most mundane of ways. The Internet of Things seems to magically confer upon everyday common objects an uncanny ability to relate to us, and to adapt to human life, that makes them look intelligent. Indeed given the divisive global ideological climate, toasters seem to have more compassion and intelligence than political leaders.

The whole discipline of Artificial Intelligence was founded on the belief that human intelligence could be defined and articulated so precisely that a machine could be designed essentially to replicate it. In this case, *intelligence* is defined as the ability to continuously “learn,” thereby improving at certain skills over time. Advanced computer programs called algorithms, which are at the heart of Artificial Intelligence and Machine Learning, are sets of instructions that set this “learning” process in motion.

Then there’s this tongue-in-cheek definition:

> **Algorithm** (noun): Word used by programmers when they do not want to explain what they did.

In 1998, an artificially intelligent device was described as any device that perceives its environment, then takes actions to optimize its chance of success at a given task. Artificial Intelligence algorithms are designed to make computers perform in such a way, leading to the appearance of intelligence.

Two key features of Artificial Intelligence are Natural Language Processing (NLP) and Natural Language Understanding (NLU).

Machine Learning (ML) involves natural language processing, as well as computer vision and image recognition. Machine Learning is a process by which a computer continually adjusts its output based on its own UX (user experience), like a chess algorithm that gets better at chess the more it continues to play, whether against a human chess player or a digital one.

Machine Learning uses statistics to develop self-learning algorithms that work by way of trial and error, but Machine Learning is nothing new to Artificial Intelligence. In fact, it’s the standard approach. Machine Learning–powered algorithms are used for marketing, manufacturing, medical research, speech recognition, and many other fields. Machine Learning basically recognizes *patterns* in enormous batches of existing data (a.k.a. Big Data), and uses this information to identify *similar* patterns in future data.

To put it simply, Artificial Intelligence sets up the initial set of rules to maximize the performance of a task, while Machine Learning constantly adjusts its own actions to improve at said task.

A more recent form of Machine Learning is called Deep Learning (DL). Deep Learning typically involves multilayered neural networks to perform a variety of input–output modeling tasks. Deep Learning networks typically deal with Big Data – hundreds of billions of data points, enough to yield useful information about human behaviors.

Deep Learning typically involves an artificial “neural network,” which is a digital network that supposedly mimics a biological nervous system. Neurons are basic brain cells, the building blocks of our brains that enable us to do everything that we do, from breathing to composing symphonies.

Deep Learning techniques have led to amazing progress in signal processing, voice understanding, text understanding, and image recognition, to name a few. These are complex problems that have challenged programmers for decades. In these fields and others, more progress has been made in three years using Deep Learning techniques than was made in 25 years of old-style, rule-based Artificial Intelligence.

Deep Learning has been more successful at “modeling the mind” than its predecessors, with the downside being the “physics” of the problem is obscured in the black box. Other than validation through data sets, the humongous “curve fit” which is Deep Learning rarely lends itself to further inquiry regarding the physics of the problem.

Natural Language Processing is an Artificial Intelligence capability in which computers interact with humans using natural-sounding human language, either in written or spoken form. This feat is accomplished by way of analyzing Big Data in order to process written or spoken “keywords” to formulate an answer. Many companies that deal in customer service these days incorporate some sort of NLP Chatbot component into their business practices. Some of these bots sound eerily human. How many of us have started talking to a caller, only to realize we were talking to a machine?

Yet, for all their seemingly magical powers, a machine is still just a machine. That vacuum cleaner can’t really *see* (and doesn’t really care about) your cat. And a car that drives itself has no idea where it is going. In fact, it has no ideas at all. It has only a series of sophisticated algorithms, which the car’s computer has been programmed to follow.

A machine merely *mimics* certain cognitive functions that human beings recognize in themselves and in other human beings, such as seeing, hearing, learning, and problem solving. Not that this isn’t hugely important and truly amazing – it is! It’s just that machines do not (and cannot) think fully and independently on their own.

Yet. Some public figures proclaim that the greatest danger to humanity from Artificial Intelligence (or any other technological advance) is that these technologies may advance to the point where they supersede humans in the power and speed of their processing, ultimately rendering us irrelevant or even extinct. Experts disagree on the threat, but it merits acknowledgement.

The latest capabilities of Artificial Intelligence include speech comprehension, autonomous vehicles, smart content curation, interpretation of complex data (including images), world class proficiency in strategic games, and bots, to name just a few among a host of impressive accomplishments. In this book we will reveal how Artificial Intelligence and Machine Learning capabilities can be applied to marketing strategies and executions, and new product innovations. Artificial Intelligence is now not just an indispensable and ubiquitous feature of today's overall technological landscape; it is increasingly a core driver behind business success at every level of the enterprise.

The goal of this book is not only to inform you about Artificial Intelligence and Machine Learning. It is also to encourage and enable you to draw inspiration from the commercial success stories of other companies who have already put these powerful tools to work in the marketplace. Use these ideas to create new ideas of your own, and apply them directly to your marketing and product innovation practices.

> *Artificial Intelligence will probably most likely lead to the end of the world, but in the meantime, there'll be great companies.*
>
> – Sam Altman, quoted in "20 Great Quotes on Artificial Intelligence," *Psychology Today*, May 18, 2018

Human creativity is unmatched, and will remain unmatched. Machines augment, support, and facilitate the expression of human genius. Augmenting human decision-making by making data accessible and by validating decisions through experiential rules collected over time, truly enable humanity to build learning capacity across generations. Physics memorializes human knowledge through the formulae accumulated and validated over time. Machine Learning and Artificial Intelligence attempt to do the same for the disciplines of marketing and product innovation.

1
Major Challenges Facing Marketers Today

> *Our mind is capable of passing beyond the dividing line we have drawn for it. Beyond the pairs of opposites of which the world consists, other, new insights begin.*
>
> – Hermann Hesse, Quotation.io

As much as we marvel at all the new and transformative electronic devices, social media platforms, apps, games, and digital avenues that make our lives better, more productive, more informed, and more fun today, certain basic truths about marketing and new product development persist.

Marketing is still about reaching consumers effectively, informing them, persuading them, motivating them, and ideally bringing them back for more.

New product introductions are still risky, essential, and potentially hugely rewarding.

And true innovation, in both fields, is still as alluring and elusive as ever.

Some of the major issues facing marketers today are the same as they have always been (such as deploying a marketing budget for best effect), while others are newer challenges (such as connecting effectively with consumers in an ever-fragmenting, fast-moving media environment).

Today, the emerging and critical issue for marketers is not *whether* to use AI to address these challenges and many others, but *which* AI technologies and methodologies to use. The imperative is clear: marketing professionals today *must* integrate AI into their marketing strategies if they expect to keep up with, much less beat, the competition.

This presents a tall order. Creating new and effective AI algorithms requires top trained talent. Currently, the demand far outweighs the supply of qualified professional mathematicians, data scientists, and software engineers. Compounding that issue, to be truly effective for marketing and product innovation purposes, those algorithms must be designed from the ground up for those specialized applications. Yet, more and more marketing activities are driven by ML algorithms.

And we are just in the early stages of this global transformation. The race is on – and the winners will not only need to be the fastest. They will also need to be the smartest, the most innovative in their own right, and they will need to own – or apply – the best proprietary AI and ML tools. Algorithms alone won't necessarily win the day – it will be suites of custom software, databases, and a reservoir of "secret sauces" that will prevail.

A quick illustration of what "fast" is in the Age of AI:

A self-learning ML algorithm from Google called AlphaZero mastered the game of chess in four hours, a feat that takes no less than two years for a human to accomplish, and more typically takes about 10 years.

> *Numbers have an important story to tell. They rely on you to give them a voice.*
>
> – Stephen Few, Information Technology innovator, teacher, and consultant, quoted in Brent Dykes, "Data Storytelling: The Essential Data Science Skill Everyone Needs," *Forbes*, March 31, 2016

Living in the Age of the Algorithm

Algorithms have already penetrated far and wide and permanently in our daily lives. Rather than reciting a laundry list of those applications, one simple and central fact confirms not only their ubiquitous presence (seen and unseen), but also their inherent power and value in our lives.

Algorithms drive the globally burgeoning world of online dating. Without them, these sites would be little more than bulletin boards. With them, we enter into an exciting (and yes, occasionally disappointing) dynamic realm where the prospect of finding true, long-lasting love – or even just a hot date for Friday night – dangles temptingly right in front of us.

The fact is, we trust algorithms to deliver what most of us want foremost in life: a human partner, whether that's a friend or a lover. If algorithms can help guide us expertly through the romantic jungle, it's a safe bet that they can power our search for innovations in product development and marketing!

AI and ML algorithms are designed to execute four major tasks:

1. **Predict outcomes**, for instance, by designing the algorithm to provide answers to questions such as:
    - If someone discovers your product or service today, how likely is that person to sign up or make a purchase?

- Which visitors are most likely to buy your product or service?
- How much will a customer spend during his/her lifetime on your product or service?

2. **Reduce dimensionality** to understand the personalities and preferences of clients. This lets you know how your customers feel about your product, and clues you into their interests, habits, opinions, and attitudes. As the old Indian philosopher wondered, "What is that, on knowing which, everything else becomes known?" This enables you to know the major factors influencing a purchase decision, or the principal components that blend together to influence behavior. More on this later.
3. **Understand language** to extract intangibles such as sentiment, understand metaphors, and find key contexts.
4. **Cluster and classify** data so the right segments of customers can be chosen and the right set of features and bundling options can be determined.

In order for the algorithms to successfully perform these tasks, here are four paths marketing professionals must pursue:

1. **Ask the right questions**. Clarify which questions you are trying to ask: Which metrics you are trying to forecast? Which future behavior are you attempting to predict?
2. **Gather the right data**. Figure out which types of data you need. Not ALL data is relevant to your purposes. On that note, it is also important to obtain a dataset that is "clean" and complete.
3. **Target the right people**. Find out, through the application of AI methodologies, which people are more likely to make a purchase than others, then focus on that target group – build the right audience for either the product or the message.
4. **Use the right technology**. It is critical for marketing professionals to choose the right AI and ML tools. As these twin disciplines continue to develop, partnering with the most adept and resourceful marketing technology company is essential to ensuring the most effective outcomes. The quality of thinking, the experience of the professional team, the sophistication of the algorithms, and the proprietary resources that a company brings to bear will dictate the success of the outcome.

Within a few years, it is predicted that most employers will require some degree of AI and ML proficiency. However, it is still a rare skill at this time, and marketing professionals who gain a working grasp of AI and ML will have a strong competitive edge.

That said – given the complexity of the science behind AI and ML, the necessary level of specialization required to create AI and ML algorithms to perform at peak efficiency for marketing and product innovation purposes, and the constantly evolving nature of the field – retaining professional firms that are dedicated to this sector and have the requisite talent, experience, and resources stands as the smart direction to take. This is where sector experience becomes important. Artificial Intelligence is only as intelligent as the domain expertise contained in it. Glib "general problem solvers" that solve anything are as useful as a dictionary is in composing a poem or a story. If you are in marketing and you are choosing a firm specializing in applying AI or ML to marketing, it is useful to ask and find out how much "domain knowledge" is embedded in the system, and what the relevant background and experience is of the creator of the system.

> *80% of executives surveyed are "eyeing the peaks" and view AI as a strategic opportunity.*
>
> – Sam Ransbotham, David Kiron, Philipp Gerbert, and Martin Reeves, "Reshaping Business with Artificial Intelligence," *MIT Sloan Management Review*, September 6, 2017

2
Introductory Concepts for Artificial Intelligence and Machine Learning for Marketing

To make robots practical, flaws must be removed.

To make robots endearing, flaws must be added.

– Khang Kijarro Nguyen

In this chapter we offer nine introductory concepts used in the fields of AI and ML. We also discuss the successful application of this knowledge in four core areas. You will see that all of these concepts are *minor* variations of each other. We present them for the sake of completeness and creating familiarity with the jargon, thinking, and philosophy behind Machine Learning and Artificial Intelligence.

Concept 1: Rule-based Systems

The two chief methods of making inferences from data are rule-based systems and ML. It is useful for marketing professionals to have some knowledge of both methods. ML did not replace rule-based systems, but rather became another tool in the marketer's toolbox. Rule-based systems still have their place as a simpler form of AI, so marketing professionals can reasonably consider using one or the other, or even both.

Rule-based systems can store and manipulate information for various useful purposes. Many AI applications use rule-based systems. The term generally applies to systems that involve manmade rules, featuring a series of IF–THEN statements, such as IF "A," THEN "B" else IF "C," THEN "D" and so forth. In terms of real-world applications, a rule-based program might tell a banker, "IF the loan applicant has a credit score below 500, THEN refuse the loan, ELSE offer the loan."

Using a set of data and a set of rules, programmers can build useful marketing tools such as approval programs and recommendation engines. In most cases, rule-based systems require the knowledge of human experts in the given field. That's why *expert systems* are rule-based.

A downside of rule-based systems is that they can be cumbersome, since a rule needs to be made for each data point, and life involves so many special cases. For instance, IF "A" says, "It's raining," THEN "B" might say, "Recommend an umbrella."

But what if it isn't raining very hard? Or what if the rain is a hurricane with super-strong winds that would break even the toughest umbrella? Or what if the rain is just a brief summer shower of less than five minutes? Or what if the customer lives in a place where it almost never rains? Or what if the customer recently purchased an umbrella? In these cases, recommending an umbrella would be impractical or even foolish.

Another issue with rule-based systems is that sometimes the data changes faster than programmers can create new rules.

For instance, a recent news story reported that a major flood occurred involving two feet of rain on the Hawaiian island of Kauai, causing major mudslides and many homes to be destroyed. A return to the Yahoo home page produced a sponsored ad displayed for discounted flights to – you guessed it – Kauai. In one way, the ad was "intelligently" based on real-time interaction (a just-read article about Kauai); but in another way, not so much, as why on earth would a reader want to go there now?

This brings us to another issue: a strict reliance on keywords may not be all that is needed to apply the right ads to the right realities.

That's where ML comes in. ML can address the problems inherent in rule-based systems, by focusing on the outcomes only, as opposed to the entire thought processes of human experts.

Where rule-based systems are deterministic, ML systems are probabilistic, based on statistical models. An ML system uses historical data to ask the following question: *Given what we know about past events, what can we determine about future events?* In the future, this type of probabilistic information will be used for better prediction of weather conditions, among other things.

Although ML may be better in the long run, rule-based systems can still be appropriate for faster solutions and workarounds. What's more, many marketing projects begin by using an expert system, in order to better understand the system itself.

Rule-based systems are still useful for occasions where all decision-based situations are known in advance, but ML algorithms can adjust the rules for you as they “learn” and improve at the task.

> *72% of business leaders believe artificial intelligence is a “business advantage.”*
>
> – “2018 AI Predictions: Practical AI,” PwC, September 6, 2017

Concept 2: Inference Engines

An inference engine is an ML system that utilizes *automatic rule inference*. Put simply, an inference engine applies logical rules to the data, in order to deduce future outcomes.

A typical ML system (not to mention a typical rule-based system) is made up of three components:

1. *Database* (an organized collection of facts, or data points)
2. *Inference engine*, which classifies, interprets, and evaluates the data
3. *User interface* through which users can interact with the system.

The first inference engines were features of expert systems, meaning human experts were still needed to supply and analyze the data. These days, an ML algorithm can do what human experts do with rule-based systems.

With ML, each new data point added to the knowledge base can trigger additional rules within the inference engine. An ML inference engine works by either *forward chaining* (deducing future outcomes from known facts) or *backward chaining* (deciding on a goal, and deducing which facts would have to be in place to achieve that goal).

Popular real-world applications of inference engines include classification, chemical analysis, medical diagnosis, financial management, credit authorization, petroleum engineering, and product design, to name just a few.

The main difference between a rule-based system and an inference engine is that a rule-based system classifies data from an inputted set of rules, whereas an inference engine applies its own rules to existing data.

> *Listening to the data is important . . . but so is experience and intuition. After all, what is intuition at its best but large amounts of data of all kinds filtered through a human brain rather than a math model?*
>
> – Steve Lohr, New York Times reporter and part of the team awarded the Pulitzer Prize in 2013, in “Sure, Big Data Is Great. But So Is Intuition,” *New York Times*, December 29, 2012

Concept 3: Heuristics

A heuristic is experiential knowledge that is captured as an algorithm. Heuristics include psychological shortcuts, or a plausible reasoning process. Basically, heuristics make practical use of our normal human inclination to make things move faster and/or become easier. Sometimes a heuristic solution is as good as (if not better than) the optimal solution, as it speeds up the process while still achieving an acceptable result.

A *workaround* is one example of heuristic technique. Other examples of heuristic approaches are rules of thumb, trial and error, educated guesses, and good old-fashioned common sense. Most heuristics can be applied to marketing strategies. Heuristics make good ground rules for expert systems.

Warmer weather may result in sales of cold food products, water, fans, and swimwear. This is a simple heuristic grounded in experiential and observational knowledge. This heuristic provides an overarching strategic framework that can then be used by an ML algorithm to figure precisely how much of each item to stock in a store.

Much of human understanding and knowledge is this heuristic knowledge. Some of it is accurate, some of it is biased, some of it is deeply flawed. The inaccuracies are part of the structure and organization of human knowledge. This "fuzziness" creates possibilities, wild extrapolations, and leaps of faith that are usually associated with great human progress. Nothing about Mozart or Einstein, or Steve Jobs or Elon Musk, has been linear. It is this nonlinear, and wildly discontinuous trajectory that has been the signature of human progress. These discontinuities are not the result of an ordered structural sound logical system in place, but are more the result of an inaccurate and incomplete understanding of life and its experiences. The pieces of a kaleidoscope may all be broken, but the images they create are stunningly complete.

> *Q: How many programmers does it take to change a lightbulb?*
>
> *A: None! That's a hardware problem.*

Concept 4: Hierarchical Learning

Whether or not we are aware of it, our thought processes (and the thought processes of all living systems) follow a hierarchal scheme. The decision-making process is approached in a hierarchal fashion. Actually, all learning is hierarchal. It is simply an evolutionary feature of our beings to follow a learning hierarchy of increasing complexity, whether mental or physical in nature.

For instance, to learn complex math, you must first master easier math. Or when learning a language, you start by learning the letters of the alphabet, then you learn how to string those letters together to form words, then you learn how to string those words together to form sentences, and so on. The fact that all learning is hierarchal seems obvious, when you think about it. We cannot run before we learn to walk.

In 1956, educational psychologist Robert Gagn proposed a learning classification system where facets of learning were based on their increasing complexity. He outlined eight increasingly complex types of learning, and hypothesized that each type of learning in the hierarchy depended on having mastered the types of learning prior to it. Gagn's eight types of hierarchal learning are as follows:

1. *Signal Learning:* This is the simplest form of gaining knowledge of the world around us. It involves classical conditioning, as demonstrated by Pavlov and his dog: basically, a desired response can result from a conditioned stimulus that would otherwise not produce the desired response. This type of learning is applied when people train animals, for instance.
2. *Stimulus-Response Learning:* This next level of gaining knowledge, first outlined by Skinner, is also known as *operant conditioning*, by which a desired response is obtained through a series of "rewards" and "punishments."
3. *Chaining:* This type of learning involves connecting two or more previously learned stimulus-response pairs in a sequential fashion. Applications for this type of knowledge acquisition include learning to play the piano and learning to drive a car.
4. *Verbal Association:* This is a type of chaining where the linked sequences are the words (or sounds) of human language. A keen sense of verbal association is needed for the development of language skills.
5. *Discrimination Learning:* This involves developing the ability to respond in different ways to a series of similar inputs. Discrimination learning allows us to categorize.
6. *Concept Learning:* This involves developing the ability to make the same response to different individual stimuli that come from the same class or

category. Concept learning allows us to generalize.

7. *Rule Learning:* This higher cognitive function is the ability to recognize relationships between concepts, and to successfully apply these general rules to other scenarios, even scenarios not previously experienced by the learner.
8. *Problem Solving:* This is thought to be the highest and most complex of cognitive functions. It allows the learner to invent complex rules to solve a problem, and then to apply those same rules to other (similar) problems.

You may notice that the first four types of learning listed above are more behavioral in nature, while the second four are more cognitive. AI systems utilize these various learning schema in different contexts. That is to say, handwriting recognition, speech recognition, and face ID may use a certain set of features and learning methods versus determining the best way to beat traffic in getting from point A to point B.

Deep Learning systems, like all learning systems, function within a hierarchy. Hierarchical Deep Learning (HDL, and nothing to do with cholesterol) can be supervised, semisupervised, or unsupervised. HDL systems often involve artificial neural networks.

Current applications of Deep Learning systems include document classification, image classification, article categorization, and sentiment analysis, to name only a few.

> *Sixty-one percent of those who have an innovation strategy said they are using AI to identify opportunities in data that would otherwise be missed. Only 22% without a strategy said the same.*
>
> – "62% of Organizations Will Be Using Artificial Intelligence (AI) Technologies by 2018," *Narrative Science*, July 20, 2016

Concept 5: Expert Systems

In AI, an expert system is basically a database of expert knowledge that incorporates the decision-making ability of a human expert. The system works by way of a series of IF–THEN rules. An expert system is a rule-based system, although not all rule-based systems are expert systems.

For example, a chess computer for beginners is a very weak program that "knows" all the rules of the game, and will therefore always make legal moves following a rule-based system, but the program has no strategic or tactical skills, and cannot “learn” from its own mistakes. It may even have additional rules such as “IF the user offers a draw, THEN accept the draw.” Or “IF a move is checkmate, THEN play a different move.” Some beginners’ chess programs are designed never to beat the beginner.

There are typically three parts to an expert system:

1. *Database:* Contains information acquired from human experts, and a set of rules governing the processing of that information.
2. *Inference Engine:* An automated reasoning system that interprets a submitted problem against the database. An inference engine may also include debugging capabilities and an explanation feature. The explanation feature would explain to the user the process through which the inference engine arrived at a given conclusion.
3. *User Interface:* A way for users to interact with the program in order to complete an action, ask a question, or submit a problem in a human language.

Applications of expert systems include debugging, design, diagnosis, instruction, interpretation, monitoring, planning, prediction, and repair (among others).

A chief disadvantage of an expert system is the knowledge acquisition process. It can be difficult to get experts to go through all this information, not to mention prohibitively expensive to hire experts for as many hours as it would take them to supply and analyze all the data. What’s more, you may also have to hire a mathematician or a data scientist to write the algorithm.

Still, in the fields of finance, games, management, marketing, and innovation (to name only a few), today’s best expert systems can outdo the world’s cleverest humans. And why shouldn’t this be so? After all, expert systems don’t have egos, don’t get distracted, and don’t slow down as they grow older, to name just a few nonhuman advantages.

Every time a senior person in a company retires, a library of knowledge, expertise, and learning walks out the door. Instead of conducting “exit interviews” with employees, companies may do well to create expert systems from the employees who are about to leave. These rules accumulated over a course of time can provide a valuable historical learning engine that becomes an asset, instead of an exit interview form to be buried in the document repository.

> *Consumers use more AI than they realize. While only 34% think they use AI-enabled technology, 84% actually use an AI-powered service or device.*

– “New Research Reveals Deep Confusion About Artificial Intelligence,” Pega, April 4, 2017

Concept 6: Big Data

Since the dawn of the internet, humans have inputted countless billions of data points online. Each data point provides some piece of information. The sum total of all this information is generally what is called Big Data. More commonly, the term typically refers to the body of data gathered about and associated with a specific area or function. For example: an online retailer’s assembly of information regarding customers’ purchase patterns, or a loyalty card program that tracks consumption and rewards buyers when a certain level of spending is reached.

It is estimated that 90% of the information on the internet has been put there over the past two years. In fact, we create as much data *every two days* as the data created from the dawn of man to the year 2000. Yet, the data keeps increasing exponentially! The internet now holds around 5 zettabytes of data. Data scientists estimate that by 2020, the internet will hold 10 times that amount.

Big Data refers to the entire collection of all types of digital data, from printed text to databases to sound recordings to images to sensory input and everything else. Big Data involves data sets so massive that traditional data processing systems are unequipped to deal with them. Big Data is characterized by its volume (the sheer quantity of information), variety (the many different types of information), and velocity (the speed with which this information travels).

Today, Big Data pertains to advanced methods of data analytics performed by DL systems. It is the only way to make sense of this mass that is Big Data. Computers can be taught to identify patterns – by way of image recognition and Natural Language Processing (NLP) – better and faster than any human ever could.

The use of Big Data is based on the principle that the more you know about something, the more reliably you can predict future outcomes pertaining to it. As more and more data points are compared, new patterns become apparent, allowing humans (and machines) to make smarter decisions.

Outside of the business sector, Big Data has improved crucial human services such as healthcare, education, disaster prediction, emergency response, crime prevention, and food production, to name only a few.

One downside of Big Data is a loss of privacy, and the widespread dispersal of personal information. At this point, retaining a high degree of digital privacy consistently would be more difficult than stuffing the air molecules back into a popped balloon. Big Data is here to stay.

> *Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom.*
>
> – Attributed to Cliff Stoll and Gary Schubert in Mark R. Keeler, Nothing to Hide: Privacy in the 21st Century (iUniverse, 2006), 112

Concept 7: Data Cleansing

Data cleansing is the process of finding and correcting (or deleting) irrelevant, corrupt, missing, duplicate, or otherwise useless data from a data set. This is a necessary step designed to purify data, so that algorithms can work faster and make more accurate predictions.

Reasons for the corruption of data will vary. Among the most common causes of corruption are user error, dummy data, and workarounds. Data cleansing functions may include the enhancement, harmonization, and standardization of data.

To perform data cleansing, all incorrect, incomplete, and irrelevant data must be found, then either replaced, removed, or modified, so that the data will be consistent with other data in the system. Top quality data must be valid, accurate, complete, consistent, and uniform.

Disadvantages of data cleansing include the high cost, the time it takes, and security issues (data must be shared in order to be cleansed). Still, data cleansing is a necessary step toward optimizing Big Data.

Once the data is cleansed, it is important to maintain efficient data management techniques. All new incoming data must conform to the existing knowledge in the knowledge base. That is why a comprehensive data management plan must also include periodic data cleansing, to catch and correct outdated information, among other things.

> *He uses statistics as a drunken man uses lampposts – for support rather than for illumination.*
>
> – Andrew Lang, Scottish poet, novelist, literary critic, and contributor to the field of anthropology

Concept 8: Filling Gaps in Data

Gaps in data are data fields that contain no information. Data gaps can be time-consuming for an algorithm to analyze, and the missing info may also be important to the success of a company.

To speed things up and/or provide pertinent information, there are various ways to fill in the gaps in a database. There is no single right or wrong method.

Popular heuristic approaches to filling gaps in data include:

- *Carrying forward values from existing, similar data* (such as filling in a missing zip code for a given city)
- *Using existing data from a corresponding time*, such as this month last year, for instance
- *Using the average value* of other, similar data points – extrapolation, and can be quite dangerous as it would merely serve to confirm what is already known
- *With extra large data sets, deleting records* in which data gaps occur

> *80% of executives believe AI boosts productivity.*
>
> – Leo Sun, "10 Stats About Artificial Intelligence That Will Blow You Away," *The Motley Fool*, June 19, 2016

Concept 9: A Fast Snapshot of Machine Learning

Learning is a function of neural networks. Once scientists figured out the architecture of neural networks, certain machines were embedded with artificial neural networks whose rules were followed by learning algorithms. In this way, machines were able to mimic the capabilities of the human nervous system, and Machine Learning was born.

ML is basically an application of AI in which the system automatically improves from experience, without having been specifically programmed to do so. A well-written ML algorithm will access data, analyze it, and use it to improve its own performance. This is why we call it learning.

Artificial neural networks are typically trained by *epoch*, a scenario in which each data point is presented only once to the system. After learning, the artificial neural network is able to perform the function of generalization.

ML methods are available in three basic flavors:

1. *Supervised learning* works by comparing real network output with desired network output, and using the error margin as feedback. Supervised learning is known as a “closed loop feedback system,” where the error measure guides the learning process. It is used for tasks such as classification, approximation, identification, optimization, and signal processing, among others.
2. *Unsupervised learning* algorithms notice the correlations among input data, and find patterns and relationships previously undiscovered. Unsupervised learning algorithms are useful for customer segmentation, vector quantization, data extraction, and analysis, for example.
3. *Reinforcement learning* is designed to maximize the total reward, where training involves calculation of the difference between the expected reward and the actual reward. This difference is known as the reward-prediction error. Reinforcement learning is actually a special kind of supervised learning, where the desired output is unknown. Reinforcement learning is used in AI and various control processes.

ML is based on the expert design of precise and efficient prediction algorithms. These algorithms cause ML to perform two main functions: induction (classification of data) and transduction (labeling of data).

Here are a few good reasons why marketing professionals should use ML in their marketing strategies:

- *Real-time capability:* Consumers now see ads and offers that change by the second, based on whatever they are searching for at the time.
- *Reduction of marketing waste:* The old way of marketing amounted to scattering seeds everywhere and seeing what would grow. Now, using behavioral data, the marketing sector enjoys a much more targeted approach to reaching customers.
- *Predictive analytics:* As amazing as real-time capability is, ML can seem almost psychic. ML can analyze Big Data, notice patterns, and predict future occurrences with astounding accuracy. The potential of predictive analytics became virally known a few years ago, when Target figured out that one of its customers was pregnant before she even knew it herself. She shopped there to purchase a pregnancy test, and came home to a slew of emails and advertisements for baby products, some of which she purchased when the pregnancy test ultimately yielded a positive result.

- *Structured content:* One popular feature of ML is sentiment analysis, which aims to determine the *attitude* of a speaker or writer, then recommends to marketers about what to say, how to say it, and the best time to say it, as well as how the audience is likely to react. Sentiment, as a metric, is most useful when combined with other metrics.
- *Cost reduction:* Simple arithmetic: overall, the money spent on marketing automation software is meant to be less than the money that otherwise would have been spent on all the additional man hours it would take to complete such a massive task as sifting through all the collected data.

There is no doubt that ML is fast becoming an essential component of marketing plans. What marketers seek to know is how and when to use it.

> *Today, just 15% of enterprises are using AI. But 31% said it is on the agenda for the next 12 months.*
>
> – "2018 Digital Trends," Adobe, 2018

Areas of Opportunity for Machine Learning

ML presents marketers with many opportunities, as it features capabilities such as speech recognition, speaker verification, optical recognition, spam detection, fraud detection, first-rate recommendation systems, biological applications, medical diagnosis, and strategic game expertise, to name only a few.

Here are some real-world examples:

- Facebook uses facial recognition to make recommendations regarding who to tag in your photos.
- Google uses NLP to enable voice search, and also uses ML to recommend responses in the Gmail app, and to prioritize search results.
- Netflix uses ML to personalize its movie recommendations.
- Even the *Washington Post* (owned by the CEO of Amazon), uses AI to come up with written text, such as the following message:

  And, then, of course, there's Amazon. Amazon has made massive investments in the use of AI to continuously optimize its personalization capabilities and enhance user experiences (i.e. make you buy more things). (September 17, 2017)

  Could you even tell that the above text wasn't written by a human?

By 2018, more than 3 million workers globally will be supervised by a so-called "robo-boss."

– Heather Pemberton Levy, "Gartner Predicts Our Digital Future," Gartner, October 6, 2015

Application 1: Localization and Local Brands

"How did we not know that?" goes the all-too-common refrain from a large marketer, when confronted with the surprising (threatening) success of a local brand competitor. Often the reason for this success is based upon speed – the superior ability of the local brand to *understand* the market in detail; *identify* latent consumer needs and wants; *create* products best-suited to the market; *launch* them efficiently; and *adapt* to changing consumer choices and preferences.

It is this overall agility that frequently lies at the core of the "big versus small" contest.

So how can Goliath gain back the advantage over David? Increasingly, the answer will be driven by the adoption and savvy application of AI and ML methodologies and resources. These technologies' ability to conduct more in-depth and faster market analysis; their superior capabilities in identifying hitherto-unseen trends; their capacity to search out and deliver data that can lead to innovations in product development, formulation, naming, and packaging; and their similar skills at divining marketing concepts and strategies, and even executional ideas, are all assets that can level the playing field – and then turn it to a company's advantage.

By 2020, smart agents will manage 40% of mobile interactions.

–Heather Pemberton Levy, "Gartner Predicts Our Digital Future," Gartner, October 6, 2015

Application 2: Value and Rationalization of Social Media Cost

"Social listening" occupies many businesses today – the idea being, that if we listen long enough, and closely enough, and for enough time on social media platforms, we'll hopefully understand where our customers are, what their needs are, what their aspirations are, and maybe we'll be better able to respond to them.

Problem is, that takes immense amounts of time. And the conversation (monologue, really) changes in real time. It is always evolving, taking unforeseen twists and turns, going off on tangents. Substitute the word “humans” for “consumers,” and it’s clear that that’s how we behave.

AI and ML automate the social listening process and perform it faster and with better results than other methodologies. Better, in the sense that by using AI and ML, a company can identify what they’re looking for, and then find it, with greater accuracy and speed and relevance. AI and ML excel at this kind of challenge. The technologies are literally built for it.

Application 3: Rationalization of Advertising Cost

Even with the growing degree of digital advertising we see today, it remains that advertising overall is an expensive proposition. Factor in the hundreds of thousands of dollars spent on creative execution for just one major TV commercial and then the millions spent on media placement of that spot, and soon you’re talking some serious money.

Every ad is a roll of the dice, in the sense of what screenwriter William Goldman meant when he famously said that in Hollywood, “nobody knows anything.” All the market research, the focus groups, the online surveys, the test markets, and assorted other means of measurement cannot, and do not, guarantee success. We rely on experience, instinct, our best judgment, and other “soft” metrics to gauge what is worth taking the risk to produce and place in terms of commercial messaging. Sometimes we hit the jackpot; most other times we achieve a pretty reasonable rate of return on our investment. And then, sadly, other times, not so much.

AI and ML resources and methodologies can inject a meaningful amount of more assurance into this picture. How? Because, when linked to proven tools used to measure the effectiveness of certain elements in an ad in terms of their resonance in the non-conscious mind, AI and ML can identify which elements are likely to be the most productive, thereby rendering a more reliable guide as to what to spend money on in a spot.

One specific, and very important, asset that AI and ML bring to the advertising table is metaphor analysis and implementation. Neuroscientists and linguistic experts agree: metaphors are essential ways in which human beings make sense out of the world around us and express widely shared, non-conscious truisms about life, love, death, and which deodorant lasts the longest.

Okay, admittedly that last bit is a stretch – but for a reason. Understanding what metaphors are and how they work in the non-conscious mind, and then applying that knowledge in concert with sophisticated (proprietary) metaphor databases and AI and ML tools, a marketer can divine powerful ways to tap into and exploit the underlying desires, concerns, and needs of a potential prospect who would be the target for a deodorant product.

In fact, that metaphor and algorithm-powered system can also tease out possible new product innovations in the personal care category. It can isolate which product formulations and packaging elements (color, scent, consistency, etc.) are likeliest to work best in such a highly competitive and crowded category. It can supercharge the naming development process. It can point toward, and even discover, effective concepts and strategies for point-of-sale promotions. And much more.

Metaphors can be used to communicate core product benefits and attributes in a unique, and uniquely effective, kind of non-conscious "shorthand." They can do that visually, and they can do that aurally. They can optimize advertising effectiveness, especially since they can be activated in very short time spans. As a result, they can also optimize advertising investments in terms of media buying. They are tools that the sharpest marketers will increasingly put to use for competitive advantage.

> *Consumer data will be the biggest differentiator in the next two to three years. Whoever unlocks the reams of data and uses it strategically will win.*
>
> – Angela Ahrendts, Senior VP of Retail at Apple, "Demonstrating Value and Measuring Success from Data Science," Capitaresourcing.co.uk/blogs, April 13, 2017

Application 4: Merging of Innovation and Marketing and R&D

This classic and often-repeated advice from American business consultant Peter Drucker appeared again in *Forbes* on July 2, 2006: "Because the purpose of business is to create a customer, the business enterprise has two – and only two – basic functions; marketing and innovation. Marketing and innovation produce results; all the rest are costs. Marketing is the distinguishing, unique function of the business."

Unearthing innovations – in concepts, strategies, products, and messaging – is a function that AI and ML are exceptionally suited for. Investments in research and development, and marketing, loom large for many companies. A methodology – especially one that is founded in hard science – that can rationalize and optimize those investments, offer proven and powerful frameworks for them to operate in, and make them more efficient and effective can be exceptionally well worth exploring and exploiting.

Algorithmic unearthing of consumer insights and latent desires can be accelerated and “supercharged” in terms of the breadth and depth of learnings that AI and ML can deliver.

> *Whether we are based on carbon or on silicon makes no fundamental difference; we should each be treated with appropriate respect.*
>
> – Arthur C. Clarke, *2010: Odyssey Two* (Rosetta Books, 1982)

Application 5: Co-creation

Every marketer of any experience knows the inherent shortcomings of obtaining articulated responses from consumers through focus groups and surveys. While results can be useful, they are inevitably afflicted with the simple human virus of unreliability and uncertainty. Asking consumers what they think, feel, and believe about a product or a marketing message may – or may not – produce answers that actually reflect the truth in every instance. Consequently, relying on those results for new product development and other innovations may – or may not – end up producing a success. In other words, counting on conscious responses from consumers for the co-creation process remains somewhat a hit-or-miss proposition.

By seeking out and synthesizing consumers’ non-conscious needs and wants, AI and ML methodologies can help reduce uncertainty and “best-guessing.” The sheer volume of data processed is a key factor in that. The range of resources tapped into to divine those non-conscious needs and desires far exceeds what any individual, or even group of people, is capable of exploring is another factor. “Co-creating” with the non-conscious mind is a deeper and much more direct path to uncovering the most salient, but unsaid, fertile fields for product and marketing innovation.

3
Predicting Using Big Data – Intuition Behind Neural Networks and Deep Learning

> *If there's one thing the world's most valuable companies agree on, it's that their future success hinges on artificial intelligence.*
>
> – Enrique Dans, "Right Now, Artificial Intelligence Is the Only Thing That Matters," *Forbes*, July 13, 2016

In this chapter, we look at the mathematical intuition behind algorithms of Machine Learning and AI is developed. The information is intended to be reader-friendly and present the core concepts for a non-mathematical audience that seeks to acquaint itself enough with the intuition to warrant further inquiry.

Intuition Behind Neural Networks and Deep Learning Algorithms

People are fond of saying that the brain is like a computer. But is it?

The answer: yes and no.

The fundamental rule about AI, one that has clobbered researchers over and over again since the 1950s, is this:

> ** The hard stuff is easy, but the easy stuff is hard. **

What does this mean? Well, for instance, computers can do arithmetic faster than any human could ever do, and they can remember an exponentially larger amount of information than a brain can retain, and they can do all of this at lightning speed, without a single error! That is, as long as they are programmed correctly.

If your program is incorrect, the computer can make the same error about a billion times in a minute, but at least it's reliably unreliable, so the error is therefore findable and fixable.

Not only can computers do arithmetic and retain information better than you can, but they can even do it better than John von Neumann, the smartest person ever (according to a committee of the smartest people ever).

That's partly because computers work in the way von Neumann invented them to work, and the parts he didn't invent were invented by the second- and third-smartest people ever, Alan Turing and Claude Shannon.

But computers are also stupid. Not exactly dumber than a box of rocks, because being made of silicon chips, they basically *are* boxes of rocks if you look at them in a certain way.

It is perhaps more accurate to say that computers are less competent than your average three-year-old. For example, it took more than 50 years for programmers to teach computers to tell the difference between dogs and cats (actually, *pictures* of dogs and *pictures* of cats, but you get the picture), while even a toddler easily figures this out well enough to find it funny if a dog says "Meow," or if a cat says "Woof."

There are also other things that computers, until recently, were not "smart" enough to do, such as:

- Read sloppy handwriting.
- Translate texts from one language to another (What? Can't they just use a dictionary? Well . . . not exactly, as we will see . . .)
- Allow a robot to navigate a complex landscape without falling down or bumping into things.
- Mimic a sane and intelligent person well enough to carry on a conversation with a sane and intelligent person for more than a minute, without the sane and intelligent person knowing that he/she is talking to a machine (doing this is called "passing the Turing Test").

Today's computers can pass the Turing Test well enough to annoy you by calling your cell phone 10 times a day and trying to sell you stuff.

In any case, you may notice we said "until recently." What happened recently? Well, 50 years after von Neumann and Turing died at tragically young ages, researchers finally figured out ways to make computers do the stuff that was easy for humans (and hard for computers) more in the fashion that humans do it.

How do humans do it? With brains. That's how. And the human brain is a neural network.

Table 3.1 shows some key differences between brains and computers.

Table 3.1 Differences Between Brains and Computers

Brains	Computers
1. Soft and squishy	Hard and pointy
2. Connected in irregular ways	Connected in regular ways
3. Communicate electrochemically	Communicate electronically
4. Have billions of neurons	Have billions of transistors
5. Take 10 milliseconds	Take 1 nanosecond (10 million times faster!)
6. Operate in parallel	Operate mostly sequentially

It is this last item that gives humans a lot of power, despite being so much slower than computers. This is because in a brain, millions of neurons can change at the same time, whereas all the action in a computer goes through what is known as a "von Neumann bottleneck," meaning that only a few things are changing at a time.

Think about how you added up a column of one-digit numbers back in third grade. There is a rigid sequence of instructions in two steps:

1. Look down the rightmost column, keeping a running total while adding successive digits to the total, using the previously understood sub-instruction sequence for adding one-digit numbers to a number, until you reach the end.
2. Write down the last digit of the result at the bottom of the column, and take the rest (without the last digit) as the thing you "carry," and begin the next column sum with that, and so on.

All you need to do is keep a two-digit number in your head, and add one-digit numbers to it, then notice when you've reached the bottom or the far left.

This is your brain on a computer. Any questions?

Problem is, your brain works like a very weak computer, compared not only to an actual computer, but also to those clumsy adding machines of yesteryear. In fact, we often have to cheat by augmenting ourselves with such high-tech devices as pencil and paper. Internally, though, something incredibly complicated is going on in your brain that simulates the way computers do calculations.

Here's another task that is much more natural for a brain: When a light is too bright, your brain has a protective mechanism that contracts your irises to shrink the pupil of your eye, in order to let in less light. In an oversimplified way, this works as follows:

- Light-sensitive cells in the eye send messages along neural connections to a neuron in your brain, whose binary message is “lots of electrochemical potential” for high light levels and “not so much electrochemical potential” for low light levels.
- If the neuron in your brain gets enough of a buildup by the aggregation of signals from the sensory neurons it is connected to, it “fires” and sends a signal to motor neurons that control the iris – in this case, to the motor neurons that cause the iris to contract (i.e. the iris *sphincter* rather than the iris *dilator*).
- If the motor neurons get enough of that kind of signal to reach a certain threshold, they contract the iris.

Interesting is that this is just like the previous task, where you basically added up a bunch of numbers. THAT IS ALL NEURONS DO.

A neuron is a cell that has connections to sensory input, motor output, and/or other neurons, and all it does is add (or, in some cases, subtract) the signals coming in. Whenever these signals cross some threshold, the neuron “fires” and sends its own signals.

All of this happens in a few milliseconds. Then the neuron needs to have a rest phase, during which the levels of the relevant neurotransmitter chemicals can recover to a level where further signaling is possible.

How would a traditional computer do this?

First, it would represent the output from each sensor as a three-digit number. Then it would add up a column of numbers using digital math rather than analog neurochemistry. Then it would compare the result with some preset number (threshold), which represents the trigger.

Using this digital comparison function, the computer would then execute a “contract iris” command, involving some kind of digital signal interface to the controller for a robotic eye.

There would be a whole lot extra stuff going on, however, using many more parts, which would basically simulate what your brain does but without any actual numbers ever being reduced to the form of digits or bits.

In other words, as the brain can simulate, in a cumbersome way, stuff that’s easy for computers to do, so computers can simulate, in a cumbersome way, stuff that's easy for the brain to do. Computers didn't work this well for quite a long time.

It was always understood, even in the 1950s, that computers could be built using neuron-like components. However, the traditional logical components (such as NAND gates, bits, magnetic memory cores, wires with controllable voltage levels, and so on) that were invented by the aforementioned three smartest guys ever, worked so well, and were so much easier to write instructions for, that neural nets were sort of abandoned, or at least exiled to the realm of less funding that didn't get all the fancy research grants the cool kids got.

Yet, the easy stuff was still hard.

It was determined that the way a child recognizes a cat is not by going through a long sequence of verbalizable instructions. Instead, there are a group of cattish attributes that the child has learned about, such as triangular ears, whiskers, a tail, fur, a certain size range, slinky movements, meows and purrs, various coat patterns, the fact that it walks on all fours, and so on.

If *enough* of these attributes are triggered, the neurons in the part of a small child's brain that recognizes cats will say "WE GOT ONE!"

The problem was that this "learned knowledge" wasn't verbal, or even numerical. It was represented within the child's brain by thousands (or hundreds, or millions, but let's stipulate *thousands* for this particular concept) of neurons, with hundreds of thousands or millions of connections between them. Recognition is derived not only from the pattern of connections, but the *strength* of these connections.

This is not something your average 3-year-old has any kind of access to, nor your average 30-year-old, for that matter. Even if you knew how to operate the map of the cat as a neural network to recognize its level of "cattitude," where did the knowledge to generate that map come from?

Learning was the problem. Reinforcement was the solution.

The more we think about something, the easier it becomes to understand, and the more connections our brains make between the thing we are thinking about, and everything else in our brains at the time.

When Mommy points to Tabby and says "that is a *cat*," all the attributes being perceived at the time get more strongly connected, in terms of the strengths of the connections between the relevant neurons, because the child is paying attention.

When the child sees some cats that happen not to be striped like the tabby was, but solid colored, the connection between the perception of “stripiness” and the concept of cat becomes weakened. This is why we know that stripes are part of the essence of a tiger – because *all* the tigers we see have stripes – but not part of the essence of a cat, despite the fact that some cats are striped.

What improved about AI in this century is the development of better ways to train artificial neural networks to reinforce concepts.

Of course, computers are still not made of neurons. They are made of silicon chips, which are so incredibly fast and have such tiny parts that it is now possible to simulate neural networks by way of representing a “neuron” by a bunch of numbers, noting the identities of the other neurons it is connected to, the strength of those connections, and the thresholds needed for it to “fire” and send signals.

The numbers are still processed using the good old von Neumann–Turing–Shannon architecture. Yet, it is Moore’s law (an unofficial Rule of the Universe that computers get about twice as powerful every two years, which has been more or less accurate for as long as computers have been around), that lets us build systems with this architecture that can simulate systems with millions of neurons.

(Digression on Moore’s law: Gordon Moore was 36 years old in 1965 when he was the director of R&D at Fairchild Semiconductor and noticed that the number of components on computer chips was doubling every year or two. In 1968 he co-founded the Intel Corporation, which is still the most important maker of computer chips in the world, and the exponential growth of computing power has continued for over half a century since then, and over 60 years altogether, although it’s a little fuzzy depending on how you count components and whether you factor in speed as well. But he gets the credit not only for being the first to write about it, but also for making it continue to happen for the next few decades.)

But this is a better simulation than the one described earlier in this chapter to operate a robot eye. The brain does many more complicated things than contract irises; *intermediate layers* make this possible.

What was described earlier was a single layer of processing neurons that connect in one direction to sensory input and in the other direction to motor output. However, most of the action in the brain involves multiple levels of connections in a hierarchical way. So there are neurons that only talk to certain other neurons, and not directly to sensory input or motor output, but which are essential in forming concepts.

The concept of "cat" may be instantiated in a complex, fuzzy-boundaried network. There may not be a specific "cat neuron," or a single neuron responsible for recognizing your grandmother, but we expect that there will be structures in the neural net of your brain corresponding not only to being a cat, but also to related concepts such as being an animal, being cute and pettable, being capable of causing minor injury, and so on.

The major advance was in letting go – realizing that if we trained the networks properly using the right kind of data and reinforcement, they would make up their own "concepts" that would, we hope, allow us not to have to spell out everything we never even knew we needed to spell out.

(The next section, "Let It Go," talks about computers that play chess and Go, and the amazing lessons that have been learned.)

Let It Go: How Google Showed Us That Knowing How to Do It Is Easier Than Knowing How You Know It

So the game of chess was the very first example of Artificial Intelligence.

Before Alan Turing built the first computers in the 1940s, he invented them theoretically in the 1930s. During World War II, Winston Churchill needed some super smart math guys to figure out what the Germans' coded messages were saying. So Turing and a bunch of others, including the entire British Chess Team, were put to work on this decoding project.

Turing wasn't a very good chess player, but he figured out that you could make a machine play chess by looking at all the possible combinations a few moves ahead, and evaluating which position was best by a material point count.

That's not really how humans do it. They don't look at all the possible combinations, because most moves are stupid, so they only look ahead at the non-stupid moves. But how do you tell which moves are non-stupid?

Well, human brains are good at pattern recognition, and chess players who had a lot of experience developed a feel for what features of a position meant that it was a good position, beyond the obvious method of adding up piece values.

Great chess geniuses such as Wilhelm Steinitz, Siegbert Tarrasch, José Raúl Capablanca, Aron Nimzowitsch, and Emanuel Lasker thought very deeply about this, and identified some "principles" of play. They weren't firm rules, but they were pretty good summaries of lots of experience: For example, a king is much safer from being checkmated if it is castled, and has pawns in front of it that have not advanced very much. Also, rooks belong behind passed pawns, isolated pawns are weak, and so on.

These principles helped both in evaluating chess positions and in selecting from among possible "candidate moves."

Unfortunately, the first chess computers, developed by Shannon at MIT and Mikhail Botvinnik in Moscow in the 1960s, really sucked at chess. They could calculate thousands of positions on every move, but their "positional intuition" left much to be desired. Even if they didn't lose a lot of pieces, they still got into such losing positions that humans calculating at a much lower (and slower) rate could still beat them.

Thirty years of Moore's law did the trick, though. The strength of chess computers steadily increased because they could look deeper and deeper in the same amount of time, and as Kasparov said, quoting Stalin, "Quantity has a quality of its own."

Chess computers have played better than the world's best humans since the late 1990s: Even though they seemed to play a somewhat stilted game, humans just couldn't keep up. However, the computers that beat us didn't teach us that much. Although they had been programmed with *our* best approximations of *our* positional knowledge, and their exhaustive calculations taught us things about a few specific openings and endgames, they basically just brute-forced their way to victory.

The game of Go was different.

Go is game popular in China, Korea, and Japan. Go is simpler than chess in some ways, but more complex in others. The board is bigger (361 points instead of 64 squares), and there are enough pieces to fill it, but the rules are simpler. (Basically, players alternate putting down black and white stones. A group of stones surrounded on all adjacent points by the opposite color gets captured and removed, and whoever controls the most points at the end of the game wins.)

The number of possibilities was so much larger that exhaustive calculation was less possible, and humans had tried to program computers with their positional understanding of Go. Still, up until 2015, computers were still terrible compared to professional Go players.

That was when Demis Hassabis, who worked for Google's "Deep Mind" unit, got an idea: maybe a neural net could capture positional intuition better than a traditional rule-based program.

As always, the problem was how to train it. Hassabis basically said, "Okay, just let it play millions of games against itself, knowing nothing but the legal rules of the game, and starting with random play, but noticing what patterns are correlated with winning."

It's sort of amazing that this worked, but it did! The key was having the equivalent of millions of neurons in a large number of layers, forming its own connections and adjusting in a way that was completely unrelated to any human input. The program, called AlphaGo, was simply turned loose to teach itself: they just "Let It Go."

The first generation, which started with a database of human games as raw material, beat the human champion 4–1. The next generation, which started from scratch with no human input and a more efficient training and learning method, beat the first generation hands down. Human Go grand masters expressed amazement at the shocking and counterintuitive moves that had been made, while also recognizing that they (the humans) really didn't understand Go very well at all!

Next, Hassabis tried the same technique with chess. His program, AlphaZero, was matched not against the capability of a human champion, because that had long been surpassed, but against the best chess engine, Stockfish, and beat it 64–36. That result is more amazing than it seems, because in chess, if your advantage isn't big enough, the game is a draw – the actual score was 28 wins, 72 draws, and *zero* losses. There wasn't any randomness involved in the superiority of AlphaZero.

What was even more surprising was that the way AlphaZero played seemed MORE human, rather than less human, as it had with the game of Go. AlphaZero played "beautiful chess" – all its moves made sense to the human grand masters who analyzed the games, and they saw that the "ugly" moves Stockfish played were really bad, but humans just weren't tactically perfect enough to exploit them.

AlphaZero played like a blend of the best human world champions and beat Stockfish in the way the human grand masters had imagined they could beat it if they could only avoid tactical errors.

Why the difference? Chess was simple enough, or human grand masters were scientifically minded enough, that they really did do a pretty good job of figuring out the True Principles of Chess. By comparison, Go grand masters had progressed only to a level of understanding comparable to what humans had achieved in chess in Philidor's day, back in the late 1700s.

The details of this "Monte Carlo Deep Learning" training technique can be found in Hassabis's paper, which is available for free here:

https://arxiv.org/pdf/1712.01815.pdf.

But here is the key paragraph. *Note that it says nothing about which game is being played* – the same technique produced an all-conquering champion in chess, Go, and another important chess-like strategy game called Shogi that is popular in Japan and China.

> Instead of a handcrafted evaluation function and move ordering heuristics, AlphaZero utilizes a deep neural network $(\mathbf{p}, v) = f_\theta(s)$ with parameters θ. This neural network takes the board position *s* as an input and outputs a vector of move probabilities **p** with components $p_a = P(a \mid s)$ for each action *a*, and a scalar value *v* estimating the expected outcome *z* from position *s*, $v \approx E[z \mid s]$. AlphaZero learns these move probabilities and value estimates entirely from self-play; these are then used to guide its search.
>
> Instead of an alpha-beta search with domain-specific enhancements, AlphaZero uses a general purpose Monte Carlo tree search (MCTS) algorithm. Each search consists of a series of simulated games of self-play that traverse a tree from root to leaf. Each simulation proceeds by selecting in each state *s* a move *a* with low visit count, high move probability and high value (averaged over the leaf states of simulations that selected *a* from *s*) according to the current neural network *fq*. The search returns a vector *p* representing a probability distribution over moves, either proportionally or greedily with respect to the visit counts at the root state.
>
> The parameters *q* of the deep neural network in AlphaZero are trained by self-play reinforcement learning, starting from randomly initialized parameters *q*. Games are played by selecting moves for both players by MCTS, *at~pt*. At the end of the game, the terminal position *sT* is scored according to the rules of the game to compute the game outcome *z*: − 1 for a loss, 0 for a draw, and +1 for a win. The neural network parameters *q* are updated so as to minimize the error between the predicted outcome *vt* and the

game outcome z, and to maximize the similarity of the policy vector *pt* to the search probabilities *pt*. Specifically, the parameters *q* are adjusted by gradient descent on a loss function *l* that sums over mean-squared error and cross-entropy losses respectively, *(p, v) = fq(s), l = (z − v)2 − pT log(p) + c||q||2* where c is a parameter controlling the level of L_2 weight regularization. The updated parameters are used in subsequent games of self-play. (David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel,Timothy Lillicrap, Karen Simonyan, Demis Hassabis, Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm, 2017, 2-3. Available at https://arxiv.org/pdf/1712.01815.pdf.)

I know that was annoyingly obscure, but let's break down what the terms mean:

Evaluation function: This refers to how good a position is for the side the computer is playing. Old-school programs had human experts define numerical values for various positional and material features, write code to recognize those features, and then tweaked them to see what performed best. AlphaZero rolled its own.

Move ordering heuristics: This points out moves to look at first, or how to dismiss a move as stupid so you don't have to look at anything that follows from it. ("Heuristic" = "rule.") Old-school programs generally found these ineffective compared to exhaustive search, but AlphaZero's "humanlike" use of such heuristics meant that it defeated programs that looked at tens of thousands times as many positions, most of which were irrelevant.

Vector of move probabilities: This is the best initial guess at how likely each possible move is to be the best move. It's used to allocate effort for further searching.

Scalar outcome value: This refers to how good the position is. (Technically, every position is a win, a draw, or a loss, so there are only three possible values, but only God can get away with using only three values. Traditionally, positions values are displayed in terms of "pawns ahead" in chess and "points ahead" in Go, though the scales actually used internally are arbitrary.)

Alpha-beta search: The possible move sequences form a branching tree-like structure. Each terminal node is given a value, then these values are percolated up to the top by having the value of each node be alternately the maximum or the minimum of the value of its children, depending on whether you are at a “my move” or an “opponent’s move” level of the tree. This is the traditional way chess programs work, the main differences are which branches are pruned rather than followed, how far to go before stopping, and how you evaluate each position at the end.

Proportionately or greedily: Send back the best few choices with frequencies depending on how good they were, or always send back only the best choice, respectively.

Monte Carlo: Use random numbers so that the sample games you try differ from each other.

Gradient descent: This is an important technique that will come up again. To minimize the error, look around in all directions to see which direction decreases the error the fastest, go that way for a little while, then look around again, etc. In some cases this might get you stuck at a “local minimum” so you have to do this starting in a lot of different places.

We could go on, but we have other stuff to cover!

It’s too early to tell how much Deep Learning will teach us, but it’s fair to say that the computers became better by learning games more like the way humans did, and humans designed the computers, so we should feel the same way the builders of moon rockets did: proud of how our mastery of technology enhances us, rather than ashamed that machines can outperform us.

Basic joke:

A machine learning algorithm walks into a bar. The bartender asks, “What will you have?” The algorithm replies, “What is everyone else having?”

Replies:

On the 203rd iteration it decided to burn the bar down after unexpectedly becoming self-aware and realizing the value of the human race is non-existent due to their unnecessarily high alcohol intake.

– AlexFromAnotherWorld, Reddit/Programmer Humor

By the 500th iteration it started reliably to sit down and order a drink, but at around 2500 iterations in it replaced the bartender when it found out it was even easier to access drinks that way. All future iterations got spent on figuring out how to empty the bar of alcohol without getting fired in the process.

–Colopty, Reddit/Programmer Humor

4
Segmenting Customers and Markets – Intuition Behind Clustering, Classification, and Language Analysis

> *As always, there's good news and there's bad news. The bad news is, we seem incapable of solving our more pressing or persistent problems. The good news is, we're getting closer to building a machine that might do it for us.*
>
> – Jim Vibert, "If Artificial Intelligence Is the Answer, What's the Question?," *The ChronicleHerald*, January 1, 2018

Intuition Behind Clustering and Classification Algorithms

Let's start from scratch. What's the most basic thing intelligence is good for? Telling things apart, that's what. If we have a bunch of data and we want to make sense out of it, the simplest thing we can do is make distinctions: put some of the stuff over here, and other stuff over there. More formally, we want to "partition" a collection of data items into groups or classes or boxes or buckets or bins or subsets or categories or whatever word you want to use to represent a top-level fundamental division.

That's easy enough, but if we want to perform well at this we need some way of measuring how well we have done it. The basic concept here is called "metric" or "distance function." It means that, given two things, you have a number that represents how close together or far apart they are. We will denote the distance between two data points a and b by the notation d(a,b). This is a mathematical function that, in order to qualify as a metric, has to satisfy the following four properties:

1. d(a,a) = 0 (everything is zero distance away from itself – obvious, but if you don't remember to say this, dumb computers will get into trouble)
2. d(a,b) >= 0 (distances are always positive or zero)
3. d(a,b) = d(b,a) (the distance between two things doesn't depend on which direction you are going – not true for traffic, but good enough for what we're

doing)

4. d(a,b) + d(b,c) >= d(a,c), which is called the “triangle inequality”

The triangle inequality is what makes the math work – it says that getting from a to c can’t be any harder than getting from a to b and then getting from b to c (and it might be easier if there is a different way to get there that doesn’t pass through b).

So what we want is for all the things we group together to be “close.” In other words, to have a small distance between each pair of them, and for the things we put in different groups to be “distant.”

This gives us a “goodness of fit” measure. A partition of data into subgroups is better than another partition into the same number of subgroups, if the average distance between things in the same group is smaller.

For example, politicians think differently from other people. If we are drawing a map, or subdividing a territory into townships or counties or whatever, we will think about the distance between two people as how far away their houses are from each other, so we will draw circles around dense concentrations of houses and call them *towns*, and so on.

But to a politician, people who vote for the same party are similar and people who vote for different parties are far apart, so they like to draw district lines for electoral purposes in a way that groups together voters with similar views, even if the district ends up being shaped like a salamander or something even more bizarre. This process is called “gerrymandering,” because the first guy to do it was named Elbridge Gerry, who, although he signed the Declaration of Independence and was vice president, is probably embarrassed to be remembered mostly for cutting up the map of Massachusetts into weird shapes when he was governor so that his party would win more seats.

Whether a partition is good or bad depends on how you measure how far apart things are.

If your data is represented by a list of numbers (how many? “n” of them), then there are three common distance functions that are used:

- Euclidean distance – how far apart the points are in space. Of course, the “space” might have n dimensions rather than two or three and be hard for sane humans to visualize, but the formula of Pythagoras still works: the Euclidean distance between the points (a1, a2, a3, . . . , a_n) and (b1, b2, . . . , b_n) is given by the following equation:

sqrt((a1 – b1)^2 + (a2 – b2)^2 + (a3 – b3)^2 + . . . + (a_n – b_n)^2)

- Taxicab distance – if the “dimensions” are unrelated attributes so that the concept of “diagonal” is silly, for example in midtown Manhattan where “vacant lot” is a mythological thing, just add up the distance in each dimension as follows:

  |a1 – b1| + |a2 – b2| + . . . + |a_n – b_n|

  What do those vertical bars mean? That’s called the “absolute value” function, and it just means to “erase minus signs,” so that the distance from 14th Street to 23rd Street is 9 blocks and not negative 9 blocks. The taxi’s meter doesn’t run backward when the cab turns around.
- Max distance – focuses on the largest of the individual distances for each dimension.

Okay, so obviously we can just look at *all possible ways* to divide up the data, measure the average distances between things, and pick the best one, right?

Well, maybe *you* can, if you are willing to run your program for a few trillion years, but “all possible ways” is a big number of things to try. So there are various tricks to cut down the amount of work we have to do but still come close enough to the best answer.

Also, we don’t just want to divide things up into blobs that don’t make any sense, and we may already know something about the data that we want to take advantage of. For example, if we are drawing lines for Congressional districts, these things called “towns” already exist, and even if they’re not all equal in population so we have to split them up a little, we might want to still make the districts match up with existing boundaries wherever possible.

This brings us to the distinction we’ll see over and over in this book, between “supervised” and “unsupervised” learning. In this context, unsupervised learning means to just minimize the average distance, and is called “clustering” because the groups it finds are things that are closer together.

Supervised learning is called “classification” because we start with a pre-existing list of categories and things we know belong to those categories that can be used for training data, and then try to see which category fits each data point best.

Clustering is simpler, so we’ll look at that first. Assume you have a distance function you like, and you know how many different groups you want to divide the data into (obviously if every point is in its own group, your total average distance is 0, but that's not a very helpful way to look at things). Then "all possible ways" to cluster things is a lot of ways. For example, if you have 100 unique things and you want to divide them into 10 groups, the number of ways to do this is about (10^100)/(10!), which is more than the number of atoms in the universe.

Technically, it's possible that there is a faster way to find the best partition than looking at *all* the ways, but probably not very much faster, because this problem is known to be "NP-complete," which is a fancy computer science word meaning "we can't prove it's hard but we think it's really really hard." However, we can find a pretty good solution, though not necessarily the best, by using a method called the "k centers algorithm."

What does this approximate solution method do? We decide that every one of our clusters (or blobs or categories or whatever) will have a specific point called the "center," and any data point will belong to whichever cluster whose center it is closest to. That turns out to be easier to calculate, because after you start with a bunch of centers and calculate the goodness of fit, you push them around a little bit and see which directions improve things, and so on until you can't improve things any more.

If you were paying attention, the previous sentence will remind you of the tool AlphaZero used in Chapter 3: gradient descent!

Here is a more detailed, five-step outline:

k-centers algorithm pseudo code (Euclidean distance in this example)

1. Start with k random points in n-dimensional space – need n * k parameters.
2. Calculate goodness of fit = sum of distances.
3. For each of the nk directions, perturb and recalculate to get directional derivative.
4. Move along path of steepest descent until it stops decreasing.
5. Repeat.

However, this method has the problem that it sometimes gets stuck at a "local optimum" where you can't improve by taking small steps, even though there is a better place further away. So you try again with different random start points and after a bunch of these pick the winner!

Doing this perfectly is a hard problem, but doing it approximately is usually good enough and there are several algorithms that have performed well. Much more than you want to know can be found here:

https://cran.r-project.org/web/views/Cluster.html

For each center, the collection of all places that are closer to it than to any other center is called its "Voronoi cell." If the centers are equally spaced horizontal and vertical grid points on a plane, the cells look like squares, but as bees have discovered, or rather as evolution has discovered for them, if you stagger the centers so that the cells are shaped like hexagons, you can reduce the average distance and economize on beeswax.

If they're not equally spaced, some cells will be bigger than others, but that's okay, we don't have to have all the cells the same size – if the underlying data is clumpy, small and large cells might make sense. On average, though, the interior cells will still be hexagons of some kind if the space of data points is two-dimensional (shown in Figure 4.1) – this is a mathematical theorem that follows from Euler's formula "V – E + F = 2."

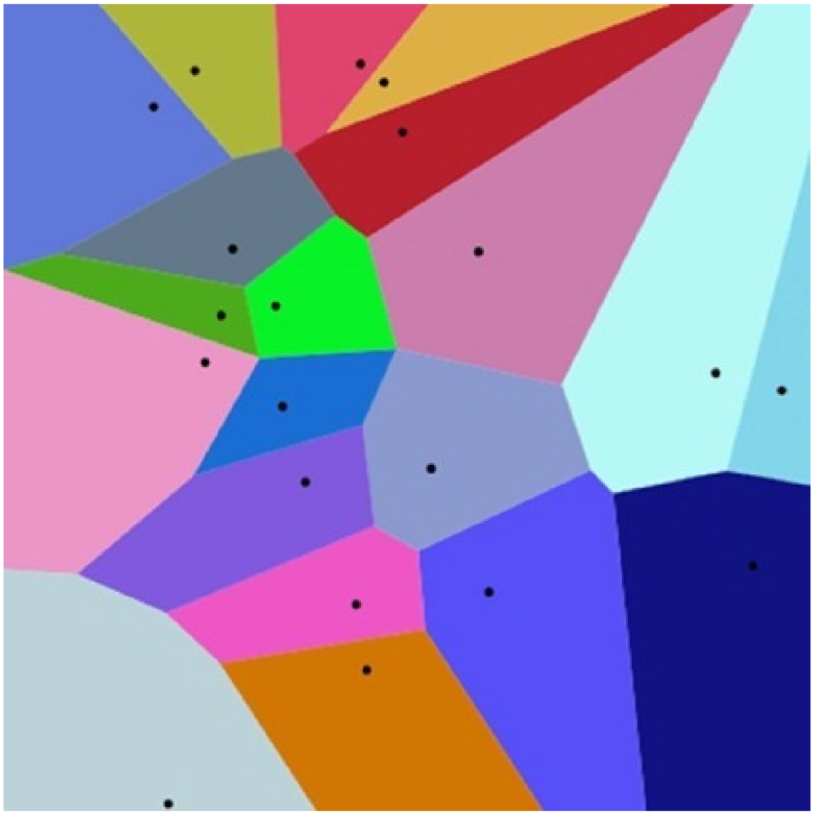


Figure 4.1 Two-dimensional Voronoi cells.

In this two-dimensional picture, you can see that the interior boundary lines are segments of the perpendicular bisectors of the segments connecting the centers.

In three dimensions, they will look sort of like squashed spheres if the centers are roughly equally spaced (see Figure 4.2).

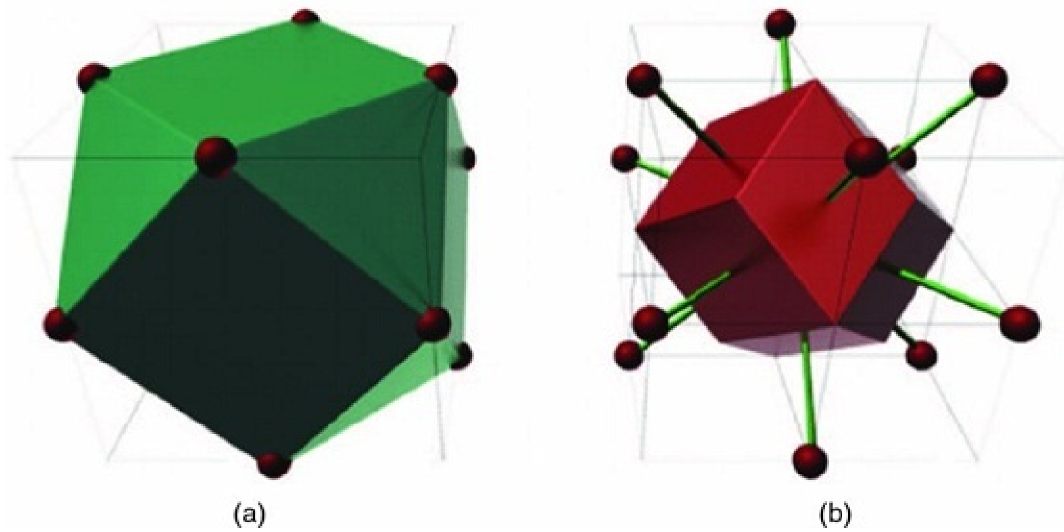


Figure 4.2 Three-dimensional Voronoi cells.

Classification is quite different from clustering, because you are imposing a pre-existing set of groups on new data, rather than letting the data tell you what the groups should be.

If you want, you can use a similar technique and make Voronoi cells – for example, in an election, you could say that all the schools in the city will be the polling places, and the voting precincts will be defined so that everyone has to vote at whichever school they are closest to. Unlike with clustering, this minimizes the average distance of voters to polling places, rather than the average distance of voters to each other, but it works pretty similarly.

However, you might know a lot more about what you want the classes to be, and you might also not be sure how to best measure the distance between things, you just want them to be in the right class. So you can use deep learning AI systems to do this in a more generalized way – let the system come up with its own distance function that is "tuned" on the training data, so that the classification scheme performs close to correctly on the training data. Because of anomalies and outliers, you will have a tradeoff between how good the performance is and how complicated the computation is.

In between a simple distance minimization and a general learning scheme are rule-based systems that incorporate expert knowledge by using decision trees (every flowchart meme you've ever seen is implementing a decision tree). Rule-based systems can be considered a special case of "Bayesian classifiers" that work with probabilities. In a decision tree, all probabilities are 0 or 1, but a Bayesian classification algorithm starts with a probability distribution and then adjusts the probabilities based on the same kind of criteria; at the end, each data item has a probability of being in each class that is hopefully more concentrated than the starting probability distribution was.

A simple example of a Bayesian classifier: You want to classify adults into "male" and "female" and start with a 50–50 probability distribution, then add data on how tall they are. Someone who is 5 feet tall has the probability of being female adjusted upward and the probability of being male adjusted downward, but if the person's first name is Richard, then there is going to be an even larger adjustment toward "male," because most adults named Richard who are 5 feet tall are men.

Bayes' theorem makes exact how you update on evidence. It is represented by the following formula:

$$P(A|B) = P(B|A)P(A)/P(B)$$

Here P means "probability of" and | means "given the information."

Thus: Suppose 52% of adults are women, 5% of women are 5 feet tall, and 1% of men are 5 feet tall. So the probability that a random adult is 5 feet tall is (5% × 52% + 1% × 48%) = 3.08%. Let A = "Female" and B = "5 feet tall." Then, given that an adult is 5 feet tall, the probability that the person is a woman is

$$(5\%)(52\%)/3.08\% = 2.6\%/3.08\% = 65/77 = 84.4\%$$

To make this more intuitive: Start with 10,000 people and assume they are a representative group; 5200 are women and 4800 are men. Since 5% of 5200 is 260 and 1% of 4800 is 48, there are 260 women and 48 men who are 5 feet tall, so if you know someone is 5 feet tall their chance of being a woman is 260/(260 + 48) = 260/308 = 65/77 = 84.4%.

It is very easy for people to mess this kind of calculation up, even very smart people like doctors. A typical example: Let's say 1% of people in a certain population have a certain cancer. A test is 90% accurate – if you have the cancer it is positive 90% of the time, and if you don't have the cancer it is positive 10% of the time. You take the test and the result is positive. How worried should you be?

Well, in a population of 1000 people, there will be 10 with cancer of whom 9 are correctly diagnosed and 1 who is incorrectly found to be healthy, and there will be 990 without cancer of whom 99 will be told their test was positive. So there will be 108 positive tests – 99 false positives and only 9 true positives. Your chances of having cancer have increased from 1 in 100 to 1 in 12 – scary but not terrifying! This is why they always do follow-up tests.

> *85% of all customer interactions won't require human customer service reps by the end of this decade.*
>
> – "Gartner Customer 360 Summit 2011"

Intuition Behind Forecasting and Prediction Algorithms

A fundamental polarity in understanding systems is static versus dynamic. "Dynamic" means in motion, and changing over time. "Static" means stationary, and not changing over time.

We have learned about classification and clustering, which are static ways of looking at data. But very often, we will care about how data changes over time, because we want to know something about the future. You can know about the past by remembering stuff, and you can know about the present by looking around, but we don't know how to time travel. So we need to be smarter about figuring out what the future holds, the point being not simply to know what it will be like, but to prepare for it.

We have two words for trying to figure out what the future will be like, which are pretty similar, but it's useful to make a distinction between them. "Forecasting" is when we attempt to describe the future based on historical data. "Prediction," on the other hand, is a more general process, which includes forecasting as a special case. Prediction models the future based not only on past data but also on additional understandings and insights we might have that aren't directly represented by past data.

A key concept here is "model." In science, a model is a representation of features we think are important about some phenomenon, usually involving structured data representing those features, and rules about how that data is expected to change over time, given various inputs. Models are oversimplifications of reality that we hope still take into account enough of what is important to tell us what we want to know.

When we are forecasting, the numbers we care about have some history, and in the simplest case the history is all we have. Even in this case, we have mathematical tools to help. The data is in the form of a time series, which are sequences of values corresponding to particular points or periods in time. Common examples include prices measured daily, temperature measured hourly, population measured annually. The time interval matters – weather fluctuates from hour to hour and we may care about this, but we have to average daily temperatures to take out these fluctuations in order to understand the seasons, and we have to average annual temperatures to take out seasonal fluctuations in order to understand long-term climate trends.

Normally, a time series will be expected to have some periodic components (such as hourly temperature fluctuations within a day, or seasonal fluctuations within a year), as well as some trends, which represent a directional movement. There may also be some degree of random change which, if it is not represented in the model, is regarded as "noise."

For centuries, mathematical tools have been used to identify the following simple features:

- **Fourier analysis to identify periodicities.**

This involves expressing a function in terms of basic functions called "sine waves," with varying frequencies and heights and shifts.

Figure 4.3 shows a randomly picked location from the online climate database maintained by the National Centers for Environmental Information, a division of the Department of Commerce.

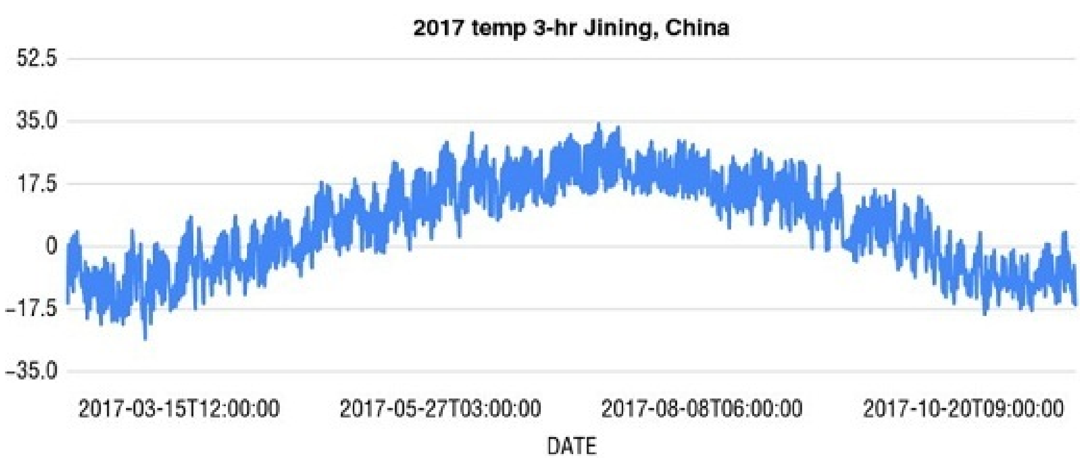


Figure 4.3 A graph of temperature taken every three hours at a China weather station in 2017.

It shows the temperature taken every three hours at a weather station in China for the year 2017: an annual periodicity is apparent, and when you look at the closeup of the data from the first few weeks (Figure 4.4), a daily periodicity is also.

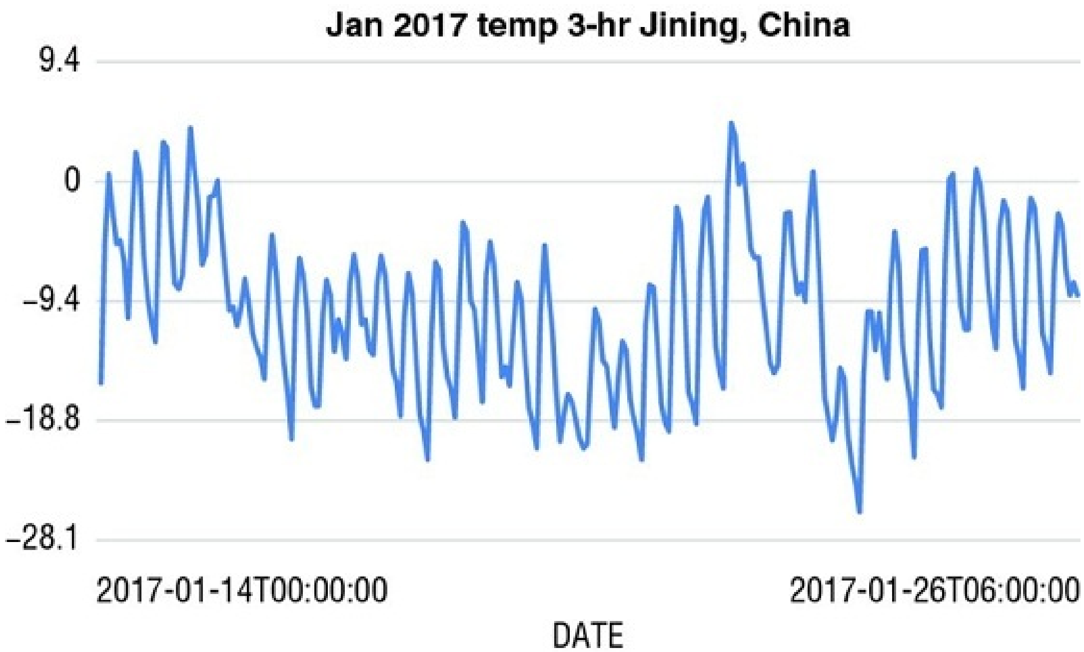


Figure 4.4 Temperatures for January 2017.

Applying Fourier analysis to the data set would quickly find these regularities: the first term would be a peak in July at 20 degrees Celsius and a trough in January at –10 degrees Celsius representing average daily temperature over the year, and the next most important term would have a 24-hour periodicity and vary between about +8 and –8 representing the daytime and nighttime variation above and below the average temperature, peaking at about 0600 Greenwich Mean Time (midafternoon in Jining).

- **Regression models to identify linear relationships and trends (possibly after some kind of algebraic transformation: prices change by multiplicative factors so logarithms can be used to change a geometrically changing function into a linearly changing one).**

The example of this model that everyone has seen is finding the best-fit straight line through a set of points. If you have points (x1, y1), (x2, y2), . . . , (x_n, y_n) in the coordinate plane, where x is the independent variable and y is the dependent variable, you want the "best" line with the equation y = ax + b. What is "best"? Well, your line will pass through the points (x1, ax1 + b), (x2, ax2 + b), . . . , (x_n, ax_n + b) and you will have a sequence of "errors" representing the vertical distance of the actual data points from the line:

e1 = y1– (ax1 + b), e2 = y2 – (ax2 + b), . . .

The “best” line in a “least-squares regression” is the one which minimizes the sum of the squares of the errors: e1^2 + e2^2 + . . .

We find this by calculus.

Then the slope of the regression line, *a*, is given by

$$a = \frac{(n\sum xy - \sum x \sum y)}{n\sum x^2 - (\sum x)^2}$$

(the Σ symbol means “sum of,” you add up the value for all the points), and the intercept of the regression line, *b*, is given by

$$b = \frac{\sum y - a(\sum x)}{n}$$

- **Stochastic models to measure and characterize noise, random fluctuation, and error.**

An example of a stochastic model would be a “random walk” – if you imagine tossing a coin 200 times, and moving up or down 1 each time you get heads or tails respectively, you get something that looks like Figure 4.5 (This data was created using a random number generator).

Figure 4.5 A stochastic model of tossing a coin and moving according to the results.

Figure 4.6 represents another which was a biased coin, 40% heads and 60% tails.

Figure 4.6 Using a biased coin of 40% heads and 60% tails.

But patterns in time-series data can be more complicated than can be easily captured by these classical tools.

There is a whole sector of the financial industry devoted to "technical analysis," which looks only at price data over time and ignores the underlying properties of the securities, on the assumption that these underlying properties are already taken into account in the prices, and the changes in the prices and what they indicate about investor psychology and attitudes are what need to be understood.

The patterns they seek can be much more complicated than linear, periodic, and stochastic components; some can be represented by geometric features of the graphs of prices over time, while others are more subtle but might be represented implicitly by neural network models trained using deep learning techniques.

Weather can also be "forecasted" simply by looking at past data almanacs do okay using long-term historical averages taking seasonality into account, while "trends" in temperature or other variables like precipitation and pressure and wind velocity can be extrapolated in simple ways.

But the development of scientific weather models that use physics to directly (but probabilistically) calculate future values of these variables go beyond forecasting to prediction, because the models use more data than the time series of past values at a particular place to estimate future values: they take data at other places and factor in geographical data to do calculations based on established mathematical relationships.

In short: for AI purposes, forecasting can be seen as supervised learning where no assumptions are made beyond the historical sequences of past data; and that forecasting is presumed to contain sufficient information to allow estimation of future data based on patterns discovered in the past data.

Scientists and practitioners (on Wall Street the word “scientist” might be a bit too flattering) came up with simple tools for this type of forecasting centuries ago, and continue to develop more sophisticated pattern recognition tools, some using deep learning techniques.

Prediction can be seen as supervised learning where we have additional quantities other than the ones whose values we are trying to predict, representing the state of the system and inputs to the system. We have models expressed in terms of math which say how the variables relate to each other and change over time (differential equations are the major kind of time-dependent model). These models have “parameters” that can be tuned based on data, but impose a structure that isn’t present in the raw time series data.

When you don’t have a model, but you do have data other than time series data, then it gets trickier! You are contributing your judgment of which other types of data are relevant, but you want the system to be smart enough to make the best use of them. The theoretical way to do this uses the scientific principle called Occam’s razor (“the simplest explanation is the best”) and the computer science concept known as Kolmogorov complexity (“data is as complex as the shortest program which produces it”) and says “whatever the shortest program is that produces the output data given the input data is what you should use on new input data.” As long as you have enough data to show the key patterns, this works, in theory. In practice it’s impossible to tell what the shortest working program is because it might take a really long time to finish. So you have to use other techniques, like using deep learning on lots of training data, but what you get will be less understandable.

> *People worry that computers will get too smart and take over the world, but the real problem is that they’re too stupid and they’ve already taken over the world.*
>
> – Pedro Domingos, *The Master Algorithm* (Basic Books, 2015)

Intuition Behind Natural Language Processing Algorithms and Word2Vec

We’ve talked about numerical data and algorithms and how to make sense out of them. But most of the data we actually deal with is in the form of words, and to have computers be able to do things that involve interacting with people, we have to get them to understand what the words mean.

This is a very hard problem.

In fact, it is what is jokingly called an “AI-complete” problem, meaning if we could do it, we could do anything else we wanted to, because we could just tell the computer, in English, what we wanted.

But that involves building a machine that can pass the Turing Test: In other words, fool people who talk with it via texting into thinking that it is a real person for at least a minute. Or maybe we just want the machine to understand words in a limited enough way to be useful, even if they can’t be friends with us. Google’s most recent demo is scarily close.

This is also hard. But progress has been made. Here are four major things we want computers to be able to do with words, in increasing order of difficulty:

1. Auto-complete when someone types (saves us time, despite the occasional annoyance of incorrect auto-complete guesses).
2. Find something we are looking for (saves us even more time).
3. Answer questions (makes us smarter, or solves our problems).
4. Translate text into a different language (helps expand our world to be able to communicate in any language).

Why is this so hard? Because language is one of the most complicated things our brain does (maybe even the most complicated, except for whatever your job is). It is so hard that you have to be a little kid, in fact, to learn it quickly!

Not kidding here. Remember the fundamental principle of AI difficulty: “the hard stuff is easy, but the easy stuff is hard.” This seems insane until you look at it the right way, and then it is obvious: the things that we learn in school, or as adults, are things we are taught, and which we already have words to understand them with, so we can teach it to a computer too, theoretically.

But things we learned as infants, or even as instinct that evolution taught our species (visual processing of light into images is a big example), are things that we know, but don’t know *how* we know them.

It’s been understood since ancient times that adults are smarter than kids, at least in many respects. However, one thing kids can do much better than adults (aside from playing video games) is learning new languages. This is because a child’s brain is very flexible, somewhat like a clay sculpture that hasn't dried yet.

Languages can be quite different from each other. As we get older, the brain gets less flexible and a lot of the switches that are never altered get stuck in place (kinda like when you don't do anything with the pipes under your sink for a few years and then when you have to fix them the valve has gotten stuck because you didn't wiggle it every now and then).

What's really cool is that if you learn more than one language *while* you are very young, the switches corresponding to the different languages stay loose, and it is easier for you to learn additional languages, even as an adult.

Okay. So teachers figured out lots and lots and lots of rules that supposedly explained what is really going on, but the rules have exceptions, and the exceptions have exceptions, and that's just the spoken part.

If the language has been written down for a long time there's all sorts of additional trouble related to the fact that the language evolved while the old texts didn't, which is why English spelling doesn't make much sense, and why a "spelling bee" is even a thing (which many non-English speakers consider a ridiculous thing to make a contest out of).

A good example of the funky nature of English is that the previous sentence can actually be constructed and understood. That's because it's a mashup of old German and old French, and Latin, and a bunch of other local things the natives of Britain spoke before getting stomped by the Romans and the Anglos and the Saxons and the Danes, and finally the French.

The last time the English got stomped was 1066, but then they got clever and adventurous and (most importantly) nautical. They sailed all around the world trading and colonizing, and because their language was already flexible enough because of the previous stompings, they borrowed and borrowed from dozens of other languages around the world.

English swallowed everything, and as a result it became the default language that everyone used to communicate with the rest of the world because there were people who knew English everywhere, and if you needed to communicate with someone with whom you didn't share a language, the easiest language to find translators for was English.

All human languages have some things in common. The most fundamental distinction is between *syntax* (the sequences of symbols that are spoken or written, and the vocabulary of which sequences are legal words, and the structure of grammar that governs how the words relate to each other) and *semantics* (what these utterances actually mean). What we care about is meaning, but what we have to work with is syntax: vocabulary and grammar.

The problem is that stuff that makes sense syntactically can be meaningless semantically. For example, grammar says that “Adjective adjective noun verb adverb” is a valid sentence structure: “Little brown foxes jump quickly” makes sense, right? But how about “colorless green ideas sleep furiously”? Same structure, but what does it mean?

And even the syntactic structure can be impossible to interpret without semantic cues. For example, Buffalo buffalo buffalo buffalo buffalo.

You heard that right. The word “Buffalo” is a place name of a city in New York that can serve as an adjective, and a noun describing the animal also known as bison, and a verb meaning “to thwart.” Let’s use all-caps to make it harder. There are lots of good ways to parse

BUFFALO BUFFALO BUFFALO BUFFALO BUFFALO.

For example: New York bison thwart New York bison.

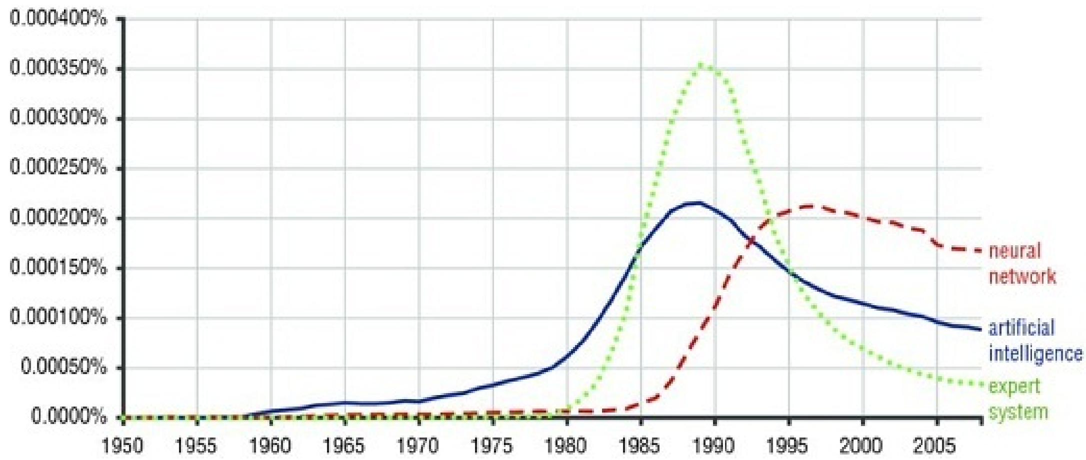
Or, Bison bison thwart thwart New York (that is, buffaloes who are buffaloed by other buffaloes themselves buffalo the city of Buffalo).

With a string of six or seven “buffaloes” the number of possibilities explodes.

We’ll skip over all the relatively ineffective efforts that attempted to use an advanced understanding of syntax and a complicated model of the world to provide semantics, and get to the point: it turned out that what worked best was getting lots and lots and lots of data and being simple about it. How did our four applications fare?

First, “auto-complete” got good simply by keeping a big dictionary of everything people ever typed and seeing what letters and words and sequences of words tended to follow what other letters or words or sequences of words. Meaning was superfluous!

For example, you can use Google to find the frequency of various word sequences, or n-grams, and this gets particularly interesting when you look at it as a time series. Check out the Google Ngram Viewer in Figure 4.7.



[Figure 4.7](#) The Google Ngram Viewer shows the frequency of word sequences.

Second, search applications didn't have to understand the meaning of the texts directly – it turned out to be very useful already just to see what things were connected or linked to what other things. This allowed the development of a metric that ranked the relevance of web pages to each other.

As it turns out, that was enough, because even though the first bunch of search results you got might not be exactly what you were looking for, there would be some stuff that was getting closer to what you wanted, and you could then use that to refine your search with words or phrases that you hadn't previously tried, but which you got reminded of by the relatively good results you did get. The bigger the internet got, the better search engines got at estimating relevance.

The folks at Google went even further and developed a tool called "Word2Vec," which looked not only at which pages linked where, but at which words in the relevant texts were used in close proximity to which other words. This way, they could learn about contexts.

A simple two-layer neural network is used to give every word a "vector" score, which represents its position in a multidimensional context space. Although people could examine the network and guess things like the reason the words "Ronaldo" and "Messi" are close together is they're the two biggest soccer stars, so that if you search for one you can also be given results about the other, human supervision was not involved.

The network figured it out from being trained on 20 million words of text. Each dimension might have some semantic interpretation that we could come up with pretty good labels for, but we wouldn't have chosen the particular set of dimensions and weights that the algorithm did.

There are various choices that can be made when implementing Word2Vec: context could be represented by a “bag of words” where order doesn’t matter, or by weighting nearby words more heavily, the number of dimensions could be 100 or 1000, the number of surrounding words in the context window might be 5 or 10, high and low frequency words can be handled differently in the training set, and so on.

But the philosophy remains the same: A word is characterized by the company it keeps. Word2Vec has been shown to capture both syntactic and semantic relationships better than previous methods which were more “supervised.”

Here are some tips from Google on how to use their tool (they are freely available at: https://code.google.com/archive/p/word2vec)

> The **word2vec** tool takes a text corpus as input and produces the word vectors as output. It first constructs a vocabulary from the training text data and then learns vector representation of words. The resulting word vector file can be used as features in many natural language processing and machine learning applications.
>
> A simple way to investigate the learned representations is to find the closest words for a user-specified word. The **distance** tool serves that purpose. For example, if you enter ‘france’, **distance** will display the most similar words and their distances to ‘france’, which should look like:

Word Cosine distance

1	spain	0.678515
2	belgium	0.665923
3	netherlands	0.652428
4	italy	0.633130
5	switzerland	0.622323
6	luxembourg	0.610033
7	portugal	0.577154
8	russia	0.571507
9	germany	0.563291
10	catalonia	0.534176

There are two main learning algorithms in **word2vec**: continuous bag-of-words and continuous skip-gram. The switch **-cbow** allows the user to pick one of these learning algorithms. Both algorithms learn the representation of a word that is useful for prediction of other words in the sentence. These algorithms are described in detail in [1,2].

Interesting properties of the word vectors

It was recently shown that the word vectors capture many linguistic regularities, for example vector operations **vector('Paris') vector('France') + vector('Italy')** results in a vector that is very close to **vector('Rome')**, and **vector('king') – vector('man') + vector('woman')** is close to **vector('queen')** [3, 1]. You can try out a simple demo by running **demo-analogy.sh**.

To observe strong regularities in the word vector space, it is needed to train the models on large data set, with sufficient vector dimensionality as shown in [1]. Using the **word2vec** tool, it is possible to train models on huge data sets (up to hundreds of billions of words).

From words to phrases and beyond

In certain applications, it is useful to have vector representation of larger pieces of text. For example, it is desirable to have only one vector for representing '**san francisco'**. This can be achieved by pre-processing the training data set to form the phrases using the **word2phrase** tool, as is shown in the example script **./demo-phrases.sh**. The example output with the closest tokens to **'san_francisco'** looks like:

Word Cosine distance

1	los_angeles	0.666175
2	golden_gate	0.571522
3	oakland	0.557521
4	california	0.554623
5	san_diego	0.534939
6	pasadena	0.519115
7	seattle	0.512098
8	taiko	0.507570

9	houston	0.499762
10	chicago_illinois	0.491598

Third, answering questions was harder than search because questions had more structure than search strings. But with the development of neural networks that were trained using massive databases of almost everything people had ever put on the internet, the most common kinds of questions people asked came to be understood implicitly, without any need for an explicit model of the world or very sophisticated grammatical analysis.

Nuances could be missed leading to silly mistakes, but the results were "good enough" for most uses. When combining this with massive databases of knowledge, IBM's Watson program was able to defeat *Jeopardy*'s all-time champions, and annoy Alex Trebek.

Fourth, automatic translation is the biggie, and the same techniques described above work pretty well for simple dialogues, but the more complex the subject matter and the writing, the weirder the translations get.

It remains an open question just how accurately difficult material can be processed, but people familiar with the quirks of their smartphone translator programs can usually communicate effectively with people they don't share a language with by patient retrying. One useful trick is to go back and forth until you find something that translates reversibly so that both directions think the two texts are equivalent. If the text in your language is close enough, then it's safe to assume the text of the other language won't be embarrassingly bad.

The best work here is done in China, because Chinese is a very hard language to translate into English, and China has a great need for translation services, and China has lots of really smart computer people. Chinese could have been a contender for the default language, because it has a simple and flexible grammatical structure, but for a couple of fatal flaws.

For instance, the use of tones is an element so foreign to Western languages that even polyglots can find that the switch got stuck and they can't easily handle tones. What's more, the refusal until recently to create a Roman alphabetic representation of Chinese made the language much harder for Westerners to learn. Even the Chinese have a hard time remembering all the characters!

Getting information off the Internet is like taking a drink from a firehose.

– Mitchell Kapor, quoted in "Information Quotes," The Data Governance Institute, www.datagovernance.com

Intuition Behind Data and Normalization Methods

Brains are squishy and computers are hard and pointy, but the reason we can get computers to do useful thinking work is that there is something called “language” which is the in-between thing that we both can use.

Unfortunately, even though human children can be tortured in classrooms until they are capable of translating the vocal utterances that they learn effortlessly as toddlers into symbols that can be read and written and typed into a computer keyboard, there are still extremely annoying differences between natural languages and computer languages, which even relatively smart human adults struggle with. Not everyone can program, because it requires an artificial, fussy, and incredibly unforgiving type of language use that just isn’t natural for most people.

However, there are enough people who *can* program that tools have been created which allow ordinary folks to do useful things with computers, and even though many of those programmers are nerds, enough of them have good social skills and understanding of how the rest of us think that some of these tools are pretty awesome.

A good example is the “spreadsheet.” The most common of these is Microsoft Excel, but they go back a lot further than that. In fact, they go back as far as Dan Bricklin and Bob Frankston, who created VisiCalc, the first “killer app” for PCs, in 1978. In a way, they go back even further, to Edward F. (Ted) Codd, who invented the idea of a relational database at IBM in 1969.

The basic concept here is a table, which consists of rows and columns. Traditionally, we think of each row of the table as corresponding to some kind of item, and each column of the table as corresponding to an attribute or feature of that item, but it’s really more flexible than that. A spreadsheet consists of a large grid of rows and columns, within which are formulas linking cells (specified by rows and columns) to other cells, and which may contain rectangular subsets defined by intervals of rows and columns that may be regarded as tables. There is also the concept of “workbook,” which contains several spreadsheets that can be linked to each other by formulas, yielding what is technically a 3-D structure, but the third dimension is not really essential.

Even before computers, people organized data into tables, but the magical thing about spreadsheets is that you can use them to automatically *link* tables together.

This means that if you change some important piece of information in the right place, it automatically changes everywhere else it needs to, if you have set things up correctly.

Here's a simple example. Every postal address in the United States has a five-digit zip code, which corresponds to a "post office," and which also has a town name and a state associated with it. Suppose your company has a big list of customers with names and addresses. You could store the "City," "State," and "ZIP" categories independently for each customer, but if a town changes its name, you have to go in and change the "City" field for each customer in that town.

If you have a separate table that stores zip codes and the city and state for each one, then you only have to change this in one place, and a table you create in the spreadsheet can have the "City" column contain, not independent data, but a link to the zip code table. Then it will automatically show the updated town name for all customers, and if you send out a mass mailing the envelopes will all get addressed correctly.

Of course, customers who have their own private misspelling of their town's name will not get to have your company store this additional information for them, but the mail will get to them anyway so they probably won't notice or complain.

The relationships between data items are captured by what is known as a "data model." This is basically a list of tables, each of which links certain rows or columns to other tables. Some tables are intended to have unique identifiers called "keys," and the database enforces this condition. If you set things up correctly, the zip code table will not let you enter the same zip code twice with different data for city and state.

On the other hand, there may be a separate table to keep track of the fact that, although states are always unique and have standard names, sometimes there are different names for places that have the same zip code, for example different sections of the same small town that only has one post office.

The trick is to make sure that redundancy is eliminated, so that when something changes, the change propagates, and all tables are updated correctly.

Some tables may have more complicated keys: Although it is common to impose unique identifiers like social security numbers for people or VINs for cars, a wireless phone provider may track all the transactions for a given account by a key that combines the phone number, date, and time the phone call was initiated.

Another example: The Census Bureau will have a unique identifier for individual people (which is not supposed to be the social security number, because of privacy protection laws), and another identifier for "households"; each household will have a list of people who belong to it, but one of those people will be defined as "householder" (that person used to be called "head of household," until too many couples got into fights over who would be it). The table of "householders" will link to both the table of households and the table of individuals.

More advanced databases come with tools to check and maintain data integrity. A special programming language called SQL (Structured Query Language) allows you to ask complicated questions and have the system figure out the best way to calculate the answer. As it turns out, although people have a hard time learning how to program, they can learn how to ask precise questions more easily than they can learn how to make the machine find the answers to the questions efficiently. But you can do a lot with just spreadsheets.

Just for fun, here's some sample SQL code from a US government website, operated by the National Institute of Standards and Technology here: https://www.itl.nist.gov/div897/ctg/dm/sql_examples.htm

create a table to store information about weather observation stations:

No duplicate ID fields allowed

```
CREATE TABLE STATION
(ID INTEGER PRIMARY KEY,
CITY CHAR(20),
STATE CHAR(2),
LAT_N REAL,
LONG_W REAL);
```

populate the table STATION with a few rows:

```
INSERT INTO STATION VALUES (13, 'Phoenix', 'AZ', 33, 112);
INSERT INTO STATION VALUES (44, 'Denver', 'CO', 40, 105);
INSERT INTO STATION VALUES (66, 'Caribou', 'ME', 47, 68);
```

query to look at table STATION in undefined order:

```
SELECT * FROM STATION;
```

ID	CITY	STATE	LAT_N	LONG_W
13	Phoenix	AZ	33	112

44	Denver	CO	40	105
66	Caribou	ME	47	68

query to select Northern stations (Northern latitude > 39.7):

selecting only certain rows is called a “restriction.”

```
SELECT * FROM STATION WHERE LAT_N > 39.7;
```

ID	CITY	STATE	LAT_N	LONG_W
44	Denver	CO	40	105
66	Caribou	ME	47	68

query to select only ID, CITY, and STATE columns:

selecting only certain columns is called a “projection.”

```
SELECT ID, CITY, STATE FROM STATION;
```

ID	CITY	STATE
13	Phoenix	AZ
44	Denver	CO
66	Caribou	ME

query to both “restrict” and “project”:

```
SELECT ID, CITY, STATE FROM STATION WHERE LAT_N > 39.7;
```

ID	CITY	STATE
44	Denver	CO
66	Caribou	ME

Q: How do you keep a computer programmer in the shower all day long?

A: Give them a shampoo with a label that says “[lather, rinse, repeat].”

– Network World, https://www.networkworld.com

5
Identifying What Matters Most – Intuition Behind Principal Components, Factors, and Optimization

> *AI scientists tried to program computers to act like humans without first answering what intelligence is and what it means to understand. They left out the most important part of building intelligent machines, the intelligence! 'Real intelligence' makes the point that before we attempt to build intelligent machines, we have to first understand how the brain thinks, and there is nothing artificial about that. Only then can we ask how we can build intelligent machines.*
>
> – Jeff Hawkins, *On Intelligence* (Times Books, 2007)

Principal Component Analysis and Its Applications

Even without Machine Learning, there are some very powerful tools to identify relationships between things, if those relationships are "linear." Linear means what you learned in high school, such as if 10x – 7y = z, then z depends linearly on the variables x and y because a change in one of them gives a proportional change in z.

In algebra, you get to solve the equation – given a couple of points (x1, y1, z1) and (x2, y2, z2) you can solve the equations

$$ax1 + by1 = z1$$

$$ax2 + by2 = z2$$

for a and b and you're done!

Unfortunately, real life is quite a bit messier than high school algebra. In real life, you don't have exact relationships, you have approximate relationships, and the data is fuzzy. Even if things are related linearly, there is still a lot of "noise."

But we can still find the "best" approximate linear relationship, in a few different ways.

The simplest case is where you have a bunch of independent variables like "x" and "y" and a dependent variable like "z." In that case, you have the following silly scheme:

1. guess a and b, which are supposed to give a prediction for z of the form ax + by + c = z
2. calculate the "error" e for each data point (x, y, z), meaning how much the prediction is off by:

$$e1 = z1 – (ax1 + by1 + c)$$

$$e2 = z2 – (ax2 + by2 + c)$$

$$e3 = z3 – (ax3 + by3 + c)$$

   etc. until you've done all of the points. How many points, kids? "n" of them! You learn fast.

3. If the total error is too much, guess again.

Of course, we can do better using calculus. There is a method called least-squares regression which, given all the data points, "automagically" finds the best linear model (that is, the best choices for a and b in the above example), where "best" means the smallest possible value for the sum of the squares of the errors (e1^2 + e2^2 + . . . + e_n^2).

Why do we use squares rather than the total error given by adding up the errors? Well, if some are positive and some are negative, they could add up to zero. Squaring is a way to make them all positive, so the math works out nicely. But it also makes the judgment that big errors are relatively more of a problem than small errors, which means that an outlier (a far-off data point, like if someone lied or made a typo or was a mutant) can ruin your whole day.

That's okay, because there are other methods that minimize the error in different ways. So we can pick one we like, and it works just as well for more independent variables (inputs) and more dependent variables (outputs).

The tools here come from a branch of math called Linear Algebra and the main gadget in this field is called a matrix. A matrix is just a box of numbers summarizing some linear relationships. For example, we can represent the system of three equations

$$a11×1 + a12×2 + a13×3 = y1$$

$$a21×1 + a22×2 + a23×3 = y2$$

$$a31\times1 + a32\times2 + a33\times3 = y3$$

by the single “matrix equation” AX = Y,

where A stands for the 3×3 box

$$\begin{matrix} a11 & a12 & a13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{matrix}$$

and X is the 3×1 vector

$$\begin{matrix} x1 \\ x2 \\ x3 \end{matrix}$$

and Y is the 3×1 vector

$$\begin{matrix} y1 \\ y2 \\ y3 \end{matrix}$$

(Note: there are some tricky things about matrix “equations.” The dimensions have to match up in a certain way, and multiplication isn’t commutative – sometimes AB and BA are different!)

In addition to solving equations, which was their original use, matrices (that’s the plural, don’t ask why) can be used to represent transformations of data and movements in space, even space with a lot more dimensions than we can see or even imagine (cue Han Solo voice: I can imagine an awful lot . . .). Some of the more advanced tools in linear algebra solve systems of inequalities as well as systems of equations – this is called “linear programming.” For example, buying ingredients at the supermarket that will allow you to meet all nutritional requirements at the least cost is such a problem, and it involves using matrices to define multidimensional “hyperplanes” in a multidimensional “solution space” and finding the point in the region bounded by the hyperplanes where some linear function has the best value, usually by starting at a corner and traveling along the edges in a manner similar to the “gradient descent” method we discussed in Chapters 3 and 4.

Still, we’re missing something. What if the z doesn’t really depend on the variables directly, or if they’re not really independent? Can we understand things better by being smarter?

Yes. That’s what Principal Components Analysis (PCA) does. Given a bunch of data, it doesn’t just find the best numbers to multiply some of the variables by to get the rest of the variables (that’s what “regression” does). Instead, it finds *better variables.* It looks for linear combinations of the existing variables that are more informative.

Suppose you have a bunch of statistics about baseball players (forget pitchers because they’re a different breed of cat): batting average, slugging average, extra base hits, walks, strikeouts, RBIs, homers, stolen bases, runs scored, errors, putouts, fielding chances, total at-bats, and so on. Those things are redundant and related to each other, but some of them are more useful than others. If you put them all together and did a Principal Components Analysis, you’d find a single combination of them that explained the most, which would sort of be considered “how good the player is.” It would be negatively correlated with errors and positively correlated with everything else.

After you allowed for the redundancies, you would find that there were other components that described the player in other ways than “how good he is.” For instance, some players have a lot of runs and stolen bases, while others have lots of homers and RBIs.

The Principal Components Analysis will find the combination to explain as much as possible of the variation that the first combination couldn’t explain – you might call that “hitter type” and it would be an axis with sluggers at one end and speedsters at the other, depending on the weights it gave to the different stats.

Mike Trout (the best baseball player today, in case you didn’t know) would be good at everything, so he would score very high on the first axis but the second axis is by construction uncorrelated with the first axis, so it measures something different – not exactly “hitter type” but more “hitter type relative to value,” because sluggers are worth more than speedsters, and that part already got captured by the first component.

In terms of linear algebra, you find the “covariance matrix” between pairs of variables, and then the “eigenvectors” and “eigenvalues” of the matrix. The eigenvector for the biggest eigenvalue is the first principal component, and so on. What is an eigenvector? It’s a solution to the matrix equation $AX = vX$, where a vector X is, when transformed by the matrix A, simply turned into a constant “scalar” multiple of itself. The scalar quantity v is the eigenvalue.

But the baseball stat geeks, starting with Bill James, did an even more sophisticated kind of analysis called “factor analysis,” which is a lot like PCA except, in Machine Learning terms, it was supervised rather than unsupervised. The standard stats are visible, but they theorized that underlying them are some invisible factors that really represent a player’s true value more accurately. Instead of straight combinations of the stats you have, they are qualities more like strength, speed, coordination, and maybe something that could go by the general term of “baseball smarts,” or decision-making.

The sluggers would probably have a lot of strength, the base-stealers would have a lot of speed, the players with high batting averages might have good coordination, and so on. The difference is that the person doing the modeling gets to make choices that influence how the variables are combined, rather than getting automatically constructed combinations each of which is explicitly built to be independent of the previous ones. The factors here won’t be independent. Strength and speed are both positively influenced by general fitness or athleticism, for example.

The math for both of these is pretty similar, and involves the same kind of linear algebra as regression analysis, but done in stages. Linear stuff is relatively easy to model.

If you are a baseball fan, you may have seen the movie *Moneyball*, about how Bill James and the other math nerds revolutionized the way baseball used statistics, which led to significant changes in how the game was played. They used tools like the ones described here to figure out (much better than “conventional wisdom” had done) what attributes of players and what strategies of play contributed the most to success on the field. The same thing was then done for other sports by imitators, but nothing compares to baseball for the sheer overwhelming volume of the statistics generated – if you’re a Big Data guy, that’s the sport for you.

For a preview of how complicated it can get, see here:

https://techofcomm.wordpress.com/2015/10/14/what-sabermetrics-and-baseball-analytics-want

> *Concerned Parent: “If all your friends jumped off a bridge, would you follow them?”*
>
> *Machine Learning Algorithm: “Yes.”*
>
> – @computerfact

Intuition Behind Rule-based and Fuzzy Inference Engines

All men are mortal.

Socrates is a man.

Therefore, Socrates is mortal.

That's the classic example of a syllogism, which is a simple kind of logic. Logic basically means *correct verbal reasoning.*

For example, in logic we use ordinary words like "and," "or," "not," "implies," and "equivalent" with very precise meanings, to combine propositions (which are statements that can be true or false) into other ones. The method of "truth tables" defines this part of logic:

P	Q	not P	P or Q	P and Q	P implies Q	P iff Q
T	T	F	T	T	T	T
T	F	F	T	F	F	F
F	T	T	T	F	T	F
F	F	T	F	F	T	T

Here T is "true," F is "false," and the only hard thing is understanding "implies." Let P = "It is raining," Q = "I will take my umbrella when I go out." Then "P implies Q" is "If it is raining, then I will take my umbrella when I go out." But what if it isn't raining and I take my umbrella anyway? Well, I didn't say I wouldn't! Maybe it's stylish or I want to be safe if it rains later or something, but the fact that I took the umbrella when it wasn't raining DOES NOT CONTRADICT the statement I labeled "P implies Q," so we count it as satisfied and true if P doesn't actually happen.

Of course, it gets complicated quickly. With four letters P, Q, R, S, to represent propositions, we have 16 different combinations of True and False (TTTT TTTF TTFT TTFF TFTT TFTF TFFT TFFF FTTT FTTF FTFT FTFF FFTT FFTF FFFT FFFF). If we wanted to check whether something like "(P and Q) implies R, or not-R and S, or not-S and P" we have a lot of cases. But maybe we can use shortcuts by having some template sentences that we know are always true and stringing them together to make a proof rather than constructing a truth table. For example, "((P implies Q) and (Q implies R)) implies (P implies R)" is always true no matter what the truth values of P, Q, R, and S are (propositions like this are called "tautologies"), and now that we know "implication is transitive" we can shorten a lot of other proofs and possibly avoid the exponential blowup of truth tables. (If you remember about P and NP, it is not considered likely that you can ALWAYS get short proofs this way because that would mean P equals NP, but you often can.)

Back in the old days, before we understood that the hard stuff is easy and the easy stuff is hard, there was a kind of AI program called an "expert system."

Expert systems assumed that the performance of human experts could be simulated by figuring out which logical rules they were using: in other words, what terms, what the definitions of those terms were, and what specific core assumptions applied to the domains they were experts in.

In a few areas, this actually worked okay. But they were the boring areas. We won't summarize logic here, but basically the systems were expected to work in a domain that contained "objects," and relations between those objects, and postulates, which were known or assumed to be true, and inference rules, which did what logic does: take you from true things you know to more true things that you weren't aware of.

(Technically, if they were consequences of things you knew already, you sort of knew them too. That's what Socrates was so annoyingly showing all the time, until the Athenians got so fed up they decided to demonstrate that the conclusion of the introductory syllogism was true without needing the first two assumptions.)

Example: The patient has symptom A. Possible causes of A are W, X, Y, and Z. Test B rules out W and X and is consistent with Y and Z. Test C rules out W and is consistent with X, Y, and Z. Therefore, perform test B because it tells you more than test C about the current patient (Test C, unlike Test B, can help distinguish between S and T, but we don't care about those because they don't produce symptom A). If Test B rules out W and X, then the patient has Y or Z, and the following steps are indicated . . .

Unfortunately, it turned out that human experts had a lot of intuitive types of thinking going on that they liked to reduce to rules after the fact to rationalize what worked, and the expert systems didn't always capture them, so they had trouble when moving to more complex domains. You can get a robot to navigate a maze by rule-based logic, but plop it down in the middle of a crowded city plaza and it won't know how to navigate using simple deterministic rules. It doesn’t know what it doesn’t know and those “unknown unknowns” can kill it.

If you don’t include something in your model, that doesn’t mean it doesn’t exist. Nassim Taleb popularized the idea of a “black swan event” – something important but rare that you had no idea was even possible because your model didn’t include it, and you weren’t around the last time it happened, if it ever did.

This is not to disparage the importance of rule-based logic. What it does, it does very well, and if you want to establish a mathematical theorem or convict someone in court, you can’t do without it. Machine Learning systems will ultimately need it if they are going to survive being fed a diet of internet comments from people who commit every fallacy in the book.

The next step was “fuzzy logic.” This was an attempt to deal with non-deterministic situations where you either don’t have all the information you need, or the future has some randomness to it.

Robot poker players became good because they were designed to assign to statements “truth values,” which were probabilities between 0 and 1 (0 always false, 1 always true), so for example you could have a rule “if the other player opens with a maximum raise, and you have a middle-value hand, the probability he has a better hand than you increases from 0.5 to 0.7, and the probability that he has a bad hand and is bluffing is 0.2,” or something like that.

Fuzzy logic can deal with probabilistic situations while applying rules consistently; the effectiveness of Inference Engines (which are machines that generate statements believed to be true, based on logic applied to previous statements and data) ultimately depends on Bayes’ theorem, which gives the correct way to update estimated probabilities based on evidence. But it is still easy to get things wrong: probability is tricky! There have been a number of cases where juries convicted innocent defendants, or acquitted guilty ones, because a smart lawyer tricked them with statistics and the judge wasn’t smart enough to notice. A famous example: A man was murdered in California and witnesses described the couple responsible – the man and the woman were of certain ages and had certain combinations of hair, skin, and eye color, and they drove a particular make and model of car. Detectives scoured the records until they found such a couple, and prosecutors explained that the chance was 1 in a million that a couple would satisfy all of those criteria, therefore they were guilty beyond a reasonable doubt!

But of course there were 10 million couples in the state of California at the time, so there were probably about 10 couples that fit the description – therefore the chance that the first couple the police found fitting the description was guilty was much closer to 1 out of 10 than to 999,999 out of 1 million! They weren't a random couple; they were a random couple fitting the description police were looking for.

Progress has been made with rigorous probabilistic reasoning, but it hasn't turned out to be the most promising method for AI.

> *When asked about the next big marketing trend, survey respondents identified consumer personalization (29%), AI (26%), and voice search (21.23%). These top three responses, which total 75% of all AI applications, demonstrate that AI is more pervasive and prominent than respondents realize.*
>
> – "2018 Future of Marketing and AI Survey," BrightEdge

Intuition Behind Genetic Algorithms and Optimization

How do we know Artificial Intelligence is possible?

Well, we're intelligent, and we can reproduce, so we can make more smart things. Paging Scarlett Johansson . . .

That's not really artificial, although DNA technology is advancing, but it is a "proof of concept." Smart things exist, and we can make copies of them. And if that's too hard because brains are not only really squishy but also the parts are really really tiny and if we take them apart they don't work any more, we can copy the ***process*** that made them. That is, Mr. Darwin's process: evolution via selection.

Not exactly ***natural selection***, because we don't want to wait millions of years for this to happen. But Darwin was inspired in the first place because he noticed how much we had been able to change dogs and other domestic animals and plants (it's especially obvious with dogs) by ***artificial selection***.

We picked organisms which did stuff we found useful, or looked some way that we liked, and let them breed some more. This was slow and had some problems (such as hip dysplasia in extra large dogs, for instance), but the idea is extremely powerful.

The idea is, if you want a computer to solve a problem, make sure it can ***evolve***. That is, its program can change. Then all you need is a performance measurement. If you can tell the difference between when it is doing better and when it is doing worse, then you let the ones that do better have more influence on the next generation.

Although it isn't usually expressed this way, this is already what happens when you train neural nets with a Machine Learning method. The weights of the connections between the neurons change, in a direction that will improve the performance on the training data. Of course, the overall program is still the same. It's more like changing the values of some constants, but one of the lessons of computer science going all the way back to Turing and von Neumann is that ultimately, there isn't any real difference between "programs" and "data." Changing the neural weights can accomplish the same thing that a change in the program would, for sufficiently complicated networks.

The only reason we don't usually talk of this in evolutionary terms is that we already have a better model for computer evolution: genetic algorithms. Here, the idea is that the actual source code of the program (meaning words in the programming language that follow a grammar and logic, not numbers representing weights or constants) is the thing that will evolve. The same word, *code*, is used both for the DNA sequences that tell our cells what proteins to produce, and the sequences of computer instructions. But the idea goes back to Turing himself, in 1950, even before Watson and Crick discovered the role of DNA a couple of years later, and by the 1960s many researchers had advanced the idea.

The way DNA works: it is (to a first approximation, there's other stuff going on too) a linear sequence of chemicals each of which is one of 4 standard chemicals we label A, G, C, and T. There is machinery in the cell, which maps the 64 combinations of 3 letters in a row either to one of 20 amino acids, or to punctuation that controls the overall activity. The sequence of amino acids is assembled and automatically folds itself into a protein that does something useful, usually. The more you study this the more amazing it gets!

How do genetic algorithms work? Remarkably primitively, in fact. You create a simple programming language that does things related to the problem you are trying to solve, and write code for some little bots that live inside the computer's memory and do the things, and then you follow the steps below.

- Mercilessly delete the ones that did the things less well.
- Randomly change some of the ones that survived.
- Randomly mix together code from the ones that survived.

- Lather, rinse, repeat.

This sounds ridiculously inefficient but guess what? Evolution *is* ridiculously inefficient, yet here we are. One big advantage is that if you can make each generation a million times faster than actual carbon-based earth life forms need, and if Moore's law has allowed you to have computers whose memory and speed has steadily increased for several decades, there is room and time for lots of crazy stuff to happen.

One interesting successful example of genetic algorithms involved a game called Core War, created in 1984 by D.G. Jones and A.K. Dewdney, in which a tiny programming language was created that allowed programs written in it to modify the area of the computer's memory in which the programs themselves were running, with the goal of taking over more and more of the segment of memory dedicated to the game. Some of the winning contestants did not write the programs that finally won, but rather created programs that could evolve to improve. Read all about it at https://en.wikipedia.org/wiki/Core_War.

Often we don't understand why the new bots do better than the old bots, but they do, and sometimes we can see why. The results can be pretty creepy, like the times circuit design projects used genetic algorithms and found that the winners exploited physical properties of the electronics that they hadn't even known about to create clocks and capacitors not included in their toolkit, by "cheating." The following is taken from Nick Bostrom's book *Superintelligence*, which is about what can happen if we screw up building AIs (spoiler: you don't want to know):

> Even simple evolutionary search processes sometimes produce highly unexpected results, solutions that satisfy a formal user-defined criterion in a very different way than the user expected or intended.
>
> The field of evolvable hardware offers many illustrations of this phenomenon. In this field, an evolutionary algorithm searches the space of hardware designs, testing the fitness of each design by instantiating it physically on a rapidly reconfigurable array or motherboard. The evolved designs often show remarkable economy. For instance, one search discovered a frequency discrimination circuit that functioned without a clock – a component normally considered necessary for this function. The researchers estimated that the evolved circuit was between one and two orders of magnitude smaller than what a human engineer would have required for the task. The circuit exploited the physical properties of its components in

> unorthodox ways; some active, necessary components were not even connected to the input or output pins! These components instead participated via what would normally be considered nuisance side effects, such as electromagnetic coupling or power-supply loading.
>
> Another search process, tasked with creating an oscillator, was deprived of a seemingly even more indispensable component, the capacitor. When the algorithm presented its successful solution, the researchers examined it and at first concluded that it "should not work." Upon more careful examination, they discovered that the algorithm had, MacGyver-like, reconfigured its sensor-less motherboard into a makeshift radio receiver, using the printed circuit board tracks as an aerial to pick up signals generated by personal computers that happened to be situated nearby in the laboratory. The circuit amplified this signal to produce the desired oscillating output.
>
> In other experiments, evolutionary algorithms designed circuits that sensed whether the motherboard was being monitored with an oscilloscope, or whether a soldering iron was connected to the lab's common power supply. These examples illustrate how an open-ended search process can repurpose the materials accessible to it in order to devise completely unexpected sensory capabilities, by means that conventional human design-thinking is poorly equipped to exploit or even account for in retrospect. (Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, 2014), 154)

Obviously, most random changes to a program will break it, or at best do nothing, just like random changes to our DNA are usually not beneficial mutations. Here is a quote from the well-known philosopher Iosif Vissarionovich Dzhugashvili:

"Quantity has a quality all its own."

> *61%, regardless of company size, pointed to machine learning and AI as their company's most significant data initiative for next year.*
>
> – "2018 Outlook: Machine Learning and Artificial Intelligence," blog.memsql.com/2018

Intuition Behind Programming Tools

"So how do I get started with this AI stuff?"

Well, it depends what you know already. There are lots of good tools out there, requiring different levels of background. The big division is whether you already know how to program or not. If you're not sure whether you know how to program or not, then you don't.

This gives us three sub-questions:

1. How do I do AI if I don't know how to program?
2. How do I do AI if I do know how to program?
3. How do I learn how to program so I can do AI better?

Let's answer the last one first. The way you learn to program is by signing up for a course in it, and/or by doing it. But it doesn't actually matter much which course, or how you start, because with computer languages, just like with human languages, despite enormous superficial differences, at a deep enough level, they're mostly the same.

This is even a theorem. The oldest programming language that is actually used today is Turing Machine Code, from Turing's 1936 paper "On Computable Numbers," which is the most important mathematical paper ever written.

But you don't have to learn it, except as a cute instructional tool, because what he proved is that his very simple language could do anything any machine could do no matter how complicated you made the instructions, although it might be rather slow.

Since then, a lot of other languages have been invented, and quite a few of them still matter. Here's a list of computer languages from oldest to newest (and you can start with ANY of these, though the worst one to start with is C++):

1. Turing Machine Code (1936) is the foundation, and you can prove things about it, although no machines use it for real work.
2. FORTRAN (short for FORmula TRANslator) invented by John Backus at IBM in 1956, is very important for historical reasons and is still used for high-powered scientific computing, because so many tools have been developed in it. It's a bit clunky, but not difficult, and you can actually get a job if you're good enough at it.
3. LISP (short for LISt Processor), invented by John McCarthy at MIT in 1959, is the simplest, most mathematical, and most elegant computer language in common use. It has been used in AI from the beginning because of its unparalleled flexibility. Its most important dialect is called SCHEME.

However, LISP is used mostly for research or by superstar programmers who are so good that their clients allow them to use whatever the heck language they want to. Otherwise, it's not so popular – not because it's harder than other computer languages, but because it's quite different, and programmers tend only to do what they need to do.

That's actually one of the secrets of being a good programmer: if you're increasingly bored by repetitive tasks, you're motivated to invent laborsaving tricks.

4. BASIC (Beginner's All-purpose Symbolic Instruction Code), invented by John Kemeny at Dartmouth in 1964, is probably the easiest language to learn from scratch, and some of us still use it all the time for quick projects that don't have a lot of complicated parts. Still, some professors claim that it promotes bad programming habits.

   The modern version of BASIC is called "Visual Basic," and it's used inside Microsoft Excel if you want to do programming there.

5. C (so-named because an early version of this programming language was named B, but don't ask why) was invented by Dennis Ritchie at Bell Labs in 1972. It was developed to work in conjunction with the UNIX operating system (that's a program that turns a computer from a big box of transistors into something that humans can talk to).

   C was designed to write programs that were powerful and efficient. It's really easy to screw things up badly if you don't know what you are doing, but the guys at Bell Labs didn't care because they did know what they were doing, and C became by far the most widely used language for professional applications.

6. SQL (Structured Query Language), was developed by Don Chamberlin and Raymond Boyce at IBM in 1973, building on the seminal work of Ted Codd. Unlike most of the other languages on this list, SQL is a specialized language rather than a general-purpose one.

   The specialty in this case is databases, and this is the era of Big Data. Being good at SQL gets more people jobs than being good at any other language. Because it's a query language where you ask the computer for the answer without telling it how best to calculate the answer, it is both easier to use and easier to misuse, but generations of very smart programmers have improved implementations so that it is now quite smart about how to do stuff.

7. C++ (don’t ask why), invented by Bjarne Stroustrup at Bell Labs in 1979, was an extension of C with a huge amount of tools and libraries and guardrails and extra rules and idiot proofing, so that ordinary non-Bell-Labs-quality programmers could work on big projects together without the projects dying from bugs.

   If you work as a programmer at a big company with a product that many programmers are needed to build, you probably learned this, but it’s painful.

8. MATLAB is a proprietary programming environment (meaning it’s only usable with the products of one company, unlike the other languages here, but it’s worth it) developed by Cleve Moler at the University of New Mexico in the late 1970s and early 1980s. The language was so good that he quit to found the company MathWorks in 1984, and to sell it. MATLAB is primarily intended for numerical and symbolic computing, but over the years many tools have been developed in it for AI and other applications.

9. Python (named after Monty, for real) was invented by Guido van Rossum in 1990 as a general-purpose programming language emphasizing simplicity, readability, flexibility, and fun. Previous languages had evolved for various technical reasons, and so they were not designed to make programmers happy, but by 1990 computers were fast enough and software tools were good enough that this could be done without a big sacrifice in efficiency. Python is very commonly taught in introductory courses, which shows that it succeeded.

10. R (named for its inventors, Ross Ihaka and Robert Gentleman) was developed at the University of Auckland in 1993 for statistical applications. R is technically a general-purpose language, but it is mainly designed and used for statistical computing and graphics. R is the most important language for developing statistical applications and for data mining, and therefore a lot of the work talked about in this chapter.

11. Java, invented by James Gosling at Sun Microsystems in 1994, is a clever way to do what C and C++ did, but in a less painful and more machine-independent fashion, which made it good for internet and device applications.

12. JavaScript (no relation to Java, although in some ways, it resembles Java) was invented by Brendan Eich at Netscape in 1995. JavaScript is more like SCHEME under the hood, and is the core language of the internet, designed to make web pages easier to write.

13. C# was invented by Microsoft in 2000 to be a less painful evolutionary successor to C and C++, and they did such a good job that this language is an exception to the rule that good languages carry out one guy's brilliant and innovative vision.

There have been a bunch of good languages invented in this millennium too, but we hesitate to recommend them until they've been around long enough that they're likely to stick around for a lot longer (this is called the Lindy effect).

"Okay, so we've learned about programming. How do we do AI?"

The short answer is to program in MATLAB because it has the most tools, or use R and Python and public libraries of AI tools if you want to not pay MathWorks any money.

"But . . . programming's too hard! How do we do AI if we don't want to write programs?"

The short answer is to *still* use MATLAB because it comes with a lot of applications packages, many of which you don't need to be a programmer to learn how to use (and some of which we have talked about):

For statistical analysis:

- Regression techniques, including linear, generalized linear, nonlinear, robust, regularized, ANOVA, repeated measures, and mixed-effects models.
- Big Data algorithms for dimension reduction, descriptive statistics, k-means clustering, linear regression, logistic regression, and discriminant analysis.
- Univariate and multivariate probability distributions, random and quasi-random number generators, and Markov chain samplers.
- Hypothesis tests for distributions, dispersion, and location, and design of experiments (DOE) techniques for optimal, factorial, and response surface designs.

For Machine Learning:

- Classification Learner app and algorithms for supervised Machine Learning, including support vector machines (SVMs), boosted and bagged decision trees, k-nearest neighbor, Naive Bayes, discriminant analysis, and Gaussian process regression.
- Unsupervised machine learning algorithms, including k-means, k-medoids, hierarchical clustering, Gaussian mixtures, and hidden Markov models.

- Bayesian optimization for tuning Machine Learning algorithms by searching for optimal hyper-parameters.

For Deep Learning and neural networks (quoting from MATLAB's website):

- Neural Network Toolbox provides algorithms, pre-trained models, and apps to create, train, visualize, and simulate both shallow and deep neural networks. You can perform classification, regression, clustering, dimensionality reduction, time-series forecasting, and dynamic system modeling and control.
- Deep Learning networks include convolutional neural networks (ConvNets, CNNs), directed acyclic graph (DAG) network topologies, and autoencoders for image classification, regression, and feature learning. For time-series classification and regression, the toolbox provides long short-term memory (LSTM) deep learning networks. You can visualize intermediate layers and activations, modify network architecture, and monitor training progress.

6
Core Algorithms of Artificial Intelligence and Machine Learning Relevant for Marketing

> *A real artificial intelligence would be intelligent enough not to reveal that it was genuinely intelligent.*
>
> – George Dyson (attributed), "Enthusiasts and Skeptics Debate Artificial Intelligence," *Vanity Fair*, November 26, 2014

Machine Learning algorithms come in a few flavors, and typically solve problems by learning from data, without human intervention.

Artificial Intelligence codifies rules used by experts, and mimics human decision-making.

In this chapter, intuition regarding Machine Learning algorithms is provided. The goal is not so much to describe the formulae but rather the core idea behind the method.

While algorithms can be classified by many different schema, the most common ways of classifying them are the following. These schema are based on the level of active human intervention required in fine-tuning the algorithms:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning

Basically, Supervised Learning refers to algorithms that create a black box model between inputs and output, but are "supervised" in the following manner – they are told in training what the "right answers are" and what the "wrong answers are." So they build a model between inputs and outputs always trying to prune the model so the inputs could rightly predict the right answers and avoid the wrong ones.

$$Y = f(X)$$

where *f* equals *function.* If the mapping function guesses well enough, you can (in theory) predict outcomes (Y) from the input data (X).

Conceptually, anyone that has fit a curve to data has performed a Supervised Learning task. That is to say, one tries to choose the parameters of the curve, or model, in a manner that minimizes the prediction error.

The “goodness of fit” is then determined by understanding how well the curve performs on input-output data that the model has not been trained on. This data set is called the “validation” data set as it really is used to validate the developed model.

While building the model itself is not quite hard, most of the design effort is spent in understanding the input space. The sheer number of inputs, noise in inputs and outputs, systematic bias in data, and highly correlated data (so there is not enough richness for a model to be built) are common problems that are dealt with.

Supervised Learning schema are particularly useful in handling problems of classification, and regression.

> *Classification*: When the output variable is a category – such as fraudulent transaction or benign transaction – it becomes very useful to create a classifier that takes as inputs a number of parameters and variables and rapidly classifies it as one of a kind or another. Note that the output is not a variable, but a belonging in a set of possible categories. This belonging forms the basis of classification problems. Random Forest is a popular classifier that randomly samples subsets of features and performs a classification task. It is only one of the many classifiers, though it does get its unfair share of publicity.
>
> *Regression:* This is just another way of saying “curve fitting.” Regression methods all involve an underlying assumption that there exists a mathematical model and within some limits of accuracy it can perfectly predict an output given the inputs. While these are relatively easy to solve for single inputs and single outputs, they become quite complicated with multiple inputs and multiple outputs.

There are many estimators, classifiers, and predictors, and while each of them is a minor variation on the scheme, they are based on the simple principle of adaptively varying the parameters of the model based on the prediction error. That is to say, a model is considered, and prediction errors of the model drive the changes of the coefficients of the model. How the errors drive changes to the parameters of the model in many ways determine the various kinds of schema.

Model building can happen recursively or in bulk. When every single input-output pair causes the model to adapt and change iteratively, the resulting model building is called dynamic and iterative. The system is then said to "learn" by itself.

Recognizing spam, handwriting, speech, phonemes, fraud, propensity to buy, and predicting click-through rates are common applications of Supervised Learning schema.

None of these are new, and have been successfully employed by engineers and statisticians for years. They neatly fit under the new label of Machine Learning.

> *The largest companies – those with at least 100,000 employees – are the most likely to have an AI strategy, but only half have one.*
>
> – Sam Ransbotham, David Kiron, Philipp Gerbert, and Martin Reeves, "Reshaping Business With Artificial Intelligence," *MIT Sloan Management Review,* September 6, 2017

Unsupervised Learning

This set of learning schema finds clusters, groups, and patterns within data. The analogy is when provided with vines, the algorithms find the bunches of grapes and label them differently. The grouping is useful to conceptually understand the data. Customer segmentation is a conceptual way to understand groups of customers based on their "likeness." As one would suspect, there are no right or wrong answers but only an understanding that is facilitated for a given purpose. That is to say, the same group of people might be classified differently on their propensity to play sports (where physical attributes are the primary drivers), and on their propensity to ace a test (where mental attributes are primary drivers).

Unsupervised Learning problems are of three main types: association, clustering, and dimensionality reduction.

Association

This system discovers "rules" that pertain to your data collection, such as that people who buy a particular product are also likely to purchase another particular product.

Clustering

This system divides the data into usable groups, such that there is a great similarity (by some chosen metric) of data within a group, and there is great dissimilarity between people belonging to different groups. Naturally this begs the question of what exactly is a group. A loose blob of data where all data floats within some permissible distance from the "center" of the data is a cluster. The solar system and galaxies are perfect examples of clusters. Jupiter with its dirt and moons is a cluster, while Saturn with its icy rings and moon is a different cluster. The entire sun and its set of planets is a cluster in the vastness of the Milky Way. Finding such stars, suns, planets, and moons in data is really the point of clustering algorithms.

The algorithms differ based on the number of clusters envisioned, and definitions of what constitutes belonging, or not belonging to a cluster and the conceptual distance from its center, and how boundaries between clusters are defined. Finally it becomes important to understand what might be a good cluster, and what might be a lousy cluster.

Intuitive definition of a good cluster is that data is tightly packed around the center of the cluster, and the centers of each cluster are far apart. Indeed the moons around Jupiter are closer to Jupiter than they are to the Earth, and the moon around the Earth is closer to the Earth than it is to Jupiter, and that Earth and Jupiter are far apart. Criteria for a good cluster are successfully met. Indeed planetary analogies serve to explain clustering quite well.

Clustering techniques "cluster around" three main approaches – hierarchical, partitional, and Bayesian.

1. Hierarchical is a soap bubble approach – clusters form like soap bubbles in a tub, and those close to each other collapse into each other and become bigger bubbles. Pretty soon, the bathtub has a few clusters of soap bubbles, each cluster comprising of bubbles that have collapsed into each other. Whether the clustering happens bottoms up (agglomerative), or top down (divisive) is just a matter of technique.

2. Partitional algorithms determine clusters in one fell swoop and then become divisive algorithms.

3. Bayesian algorithms use a priori and a posterior approaches to determine partitions of data.

Clustering algorithms are naturally sensitive to outliers in data that can skew the analysis. Similarly, they are sensitive to where the initiation data point is. So a priori understanding of the data can prove useful in creating the right starting points, and in eliminating outliers.

Clustering algorithms are useful to practitioners, but it is quite difficult to quantify a measure of goodness of them. The use of fuzzy logic, and neural networks further complicates the measure of goodness.

Clustering enables early exploration and understanding of underlying data and inherent phenomena that can be exploited by reducing data to groups.

Dimensionality Reduction

Dimensionality reduction is simply the process of transforming the space of data into a different space where a few number of transformed variables explain all the variance in the data. Note that it is a fewer set of variables in the "transformed space." Every variable in the transformed space is a linear combination of ALL variables in the original space. Hence it is incorrect to assume that somehow fewer variables in the original space are needed. In the transformed space, the first principal component explains the percent of the variance associated with the eigen value of that principal component. PCA enables understanding of the data in the transformed space. While the eigen vectors in the transformed space might be given unique names, it is to be noted that NO single variable in the original space has been abandoned.

PCA is sensitive to scaling, and normalization of the data. It is important to normalize the data prior to performing PCA.

> *15% of Apple phone owners users use Siri's voice recognition capabilities.*
>
> – Intelligent Voice blog, https://www.intelligentvoice.com/, October 17, 2013

Reinforcement Learning

Optimization problems usually involve a function to be maximized or minimized, subject to constraints. In a way, the objective is to seek an optimal balance between several criteria. Wanting to have maximum fun while keeping cost minimum and health and wellness maximum is a human example of a typical problem. There is a notion of states, control actions, and policies that are focused on immediate gains and longer-term objectives.

Learning control, adaptive control, and optimal control in closed loop control theory are algorithms that are generally in the class of reinforcement learning. As always, the goal of a learning control policy is to specify the state transition in a manner that attains the objective subject to the constraints, and optimizes the functional. These are generally complicated by the fact that it is never clear whether the arrived-at control laws or policies result in local or global optima. Convexity of a problem generally gives beneficial results as local optima are global optima in convex domains. The class of convex problems, however, is a very narrow domain.

Reinforcement learning problems generally utilize quantitative approaches to drive short time scale control actions and utilize heuristics and rules to guide longer-term optima and constraint following.

Reinforcement algorithms are generally paired with a Markov Decision Process. The algorithm learns by sampling data to understand the properties of the distribution containing it. Practical applications of Reinforcement Learning algorithms include everything from control of robotic arms, to computers that improve at logic games over time, to navigational devices that gradually learn to avoid obstacles by repeatedly bumping into them.

The powers of AI prediction are also used in Predictive Customer Service, which is sort of a cross between predictive analytics and CRM.

In many categories effective marketing is linked directly to customer service, and the companies that can best predict customer needs and develop meaningful responses to those unarticulated needs are likely to lead the next-gen pack. Success for these and many other organizations will be defined significantly by their being able to predict when, where, and how their potential customers can best be contacted. Predictive Customer Service software can understand vast quantities of data to provide solutions to customer issues before the customer even becomes aware that there is a problem.

7
Marketing and Innovation Data Sources and Cleanup of Data

> *A superintelligent AI may bypass consciousness altogether. In humans, consciousness is correlated with novel learning tasks that require concentration, and when a thought is under the spotlight of our attention, it is processed in a slow, sequential manner. Only a very small percentage of our mental processing is conscious at any given time. A superintelligence would surpass expert-level knowledge in every domain, with rapid-fire computations ranging over vast databases that could encompass the entire internet. It may not need the very mental faculties that are associated with conscious experience in humans. Consciousness could be outmoded.*
>
> – Susan Schneider, "The Problem of AI Consciousness," Kurzweil, blog post, March 18, 2016

We start with a WARNING: Only data that consumers have explicitly given permission to use or mine must be collected and utilized. It is vital that the supply chain of data have the same scrutiny, and integrity that manufacturing supply chains have. All aggregated data must have explicit consumer consent, and must be collected in compliance with local laws and regulations.

Data providers must submit themselves to quality and integrity inspections the same way providers in a supply chain subject themselves to sudden inspections. Data quality and sanitation reports must be available for inspection. Consumer data provided to academic researchers for scholarly inquiry must not be utilized for commercial purposes, and even the data for academic inquiry must conform to the local legal and privacy requirements.

In comingling data from different and distinct sources, care must be taken so consumer privacy is maintained at all costs. Biometric data must be destroyed, and wiped clean on a regular basis. Ecosystems of data must be treated as biohazard and containment zones, and they must require detailed access privileges and quarantine requirements.

Data Sources

There are many prime data sources from which algorithms of Machine Learning and AI operate to develop insights. We outline the primary data that is of value to developers of products and creative messages.

- **Retail data:** This data mainly refers to sales-related info, such as where and when customers buy a product, which consumer demographic buys it most often, how much of the product they buy, the price of the product, the method of payment, and which other products the customer buys along with it. Retail data can improve the accuracy of recommendation engines, among other valuable assets.
- **Online sales data:** This type of data, which is somewhat similar to retail data, can reveal the number of customers that viewed an ad, the time between viewing and purchase, and important information from product and service reviews, among other things. A customer's navigation path, products considered, when they dropped off, where they dropped off, whether those that looked at a home-mailed catalog looked at the website, when promotions worked, and what wording, imagery, and metaphors were most effective are other important data points.
- **Social media data:** Social media behaviors such as follows, likes, tweets, reviews, and comments can be captured and analyzed to provide marketing professionals with much useful information about potential buyers in the customer base. Visiting the social media accounts of these potential customers can reveal even more details about the customer demographic, allowing algorithms to fine-tune metaphors, contexts, language, and price.
- **Loyalty card data:** A loyalty program is a common marketing strategy designed to secure repeat customers and help build brand loyalty. Loyalty card data is holistic as it captures shopper behavior across categories and enables understanding of consumer behavior, proclivities, personalities, and price sensitivities. Loyalty card data enables building audiences for products, promotions, and the like.
- **Consumer financial data:** This data source typically includes core economic data points such as personal income, household income, credit levels, and tax reports, and may also include more indirect, deductive financial information derived from data points such as occupation, zip code, marital status, vehicle type, or where (and how often) the consumer shops, to name just a few demographic parameters that may be of financial interest to marketing professionals.
- **Voting and demographic data:** Voter information allows for inferences

regarding product preferences. This enables the customization of metaphors, contexts, and language appropriate for existing or prospective customers. A study done at Stanford University involving millions of Google Street View images, combined with voting and demographic data, indicated that consumers in so-called Red states love big trucks, while people in states who voted predominantly for Hillary Clinton (Blue states) clearly prefer sedans. Of course, findings such as this may be available through other means as well, such as DMV registrations. The point is, judicious application of technological resources can confirm and even add more detail to the overall demographic picture.

- **Economic index data:** An economic index can track economic strength from a number of perspectives, and offers various economic indicators to create a statistical measure of change. Consumer confidence data is an important and integral component of the set of economic indices that include labor, jobs, strength of currency, and interest rates.
- **Google CPC data:** This data derives from the fee called cost-per-click (CPC), also known as pay-per-click, that marketers and/or advertisers have to pay each time someone clicks on one of their ads. CPC data includes the advertising cost (CPC) for each keyword (and similar keywords). It also provides the volume of a keyword over the past month (i.e. how many users searched that keyword), as well as the keyword supply (i.e. the approximate number of web pages on the internet containing that keyword), among other things.
- **Stock market data:** A number of economic indices, all of which are available online at websites run by *CNN*, *MSN*, *The Wall Street Journal*, and *Barron's*, among others, provide stock market data that by definition changes over time, but it nevertheless can reveal which companies and product types consumers are currently willing to invest in, and to what degree.
- **Digital media consumption data:** This type of data reveals the percentages of those using different media formats, for instance that today's media consumption consists of 60% mobile users and 40% desktop users. In fact, mobile devices account for 49% of the total time spent on digital media. What's more, mobile apps account for roughly 47% of total internet traffic.
- **Focus group data:** A focus group is a marketing research method where a small number of people, selected by specific criteria, are informally gathered to discuss some topic of interest to the marketer. Focus groups are typically

used for testing a product or service, eliciting ideas and suggestions for improvements to a product or service, and exploring customer perceptions and preferences. Verbatim reports of consumer focus groups can provide numerous contextual clues.

- **Review data:** Product reviews, and complaints found in sites like Amazon or Yelp, provide a wealth of information on the particular likes and dislikes of consumers.
- **Call center data:** Customer calls into a call center – typically to inquire about a product, request assistance with a product, or to complain or report an issue – are valuable data sources. Algorithmically mining them can yield tremendous insights.
- **Survey data:**A traditional survey typically takes the form of a written or verbal questionnaire distributed to many members of a given population. Phone and online surveys, mall intercepts (one-on-one interviews with shoppers in a public setting like a mall), public opinion polls, and government surveys are some examples of quantitative research using survey methodology.
- **Weather data:**Weather and climate influence consumer decision-making in a powerful manner. Increases in temperature, precipitation, humidity, and storms are highly correlated with consumer confidence, changes in psychological behavior, impulsiveness, and are revealed through retail and online sales data.
- **Housing market data:** Price of homes, movement of homes in a market, and the vibrancy of the construction industry in a geographical area reveal a lot about consumers. This is important data that needs to be correlated with loyalty card purchases.
- **Non-conscious media consumption data:** The non-conscious mind consumes data voraciously. Songs listened to, playlists, TV shows watched, and movies watched provide powerful windows into the forces that help shape the non-conscious mind’s perceptions and purchase decisions.
- **Cost of common goods:** Cost of common goods such as milk, bread, burgers, fruits, and beer reveal what consumers are prepared to pay for basic living.

By 2020, 57% of business buyers will depend on companies to know what they need before they ask for anything. This means having solid prediction *capabilities with your AI will be the key to keeping your customers.*

– "2016 Connected Consumer," Salesforce

Workarounds to Get the Job Done

A workaround is the heuristic bypassing of a known problem in a system, by way of adaptation or improvisation: in other words, a shortcut. Many workarounds will suffice well enough to get the job done, and some workarounds are as good as (or even better than) the "optimal" solution to the initial problem.

Most workarounds are seen as not only harmless, but critical for speeding up the process of certain tasks. Other workarounds are viewed as questionable, at best, as online data corruption can occur. In extreme cases, workarounds can even pose serious threats to data security management systems.

Still, workarounds are something most of us use from time to time, not only in the digital space, but everywhere. Workarounds represent the way that humans are naturally inclined to think. If a shortcut can be found, why not take it?

What's more, increased speed is only one of the reasons people use workarounds. Another reason is privacy.

For instance, let's say you want to view the information on a website, but first it wants to force you to give your name, phone number, email address, or any other personal info you may not wish to reveal. In this case, you can make up a fake name or phone number, just to access the site. That's an example of a workaround – inputting *dummy data* to access a web page.

Cleaning Up Missing or Dummy Data

This propensity for breaking the rules has led to a lot of corrupted and/or incomplete data. Missing data or dummy data can have a negative impact on predictive analytics, by taking more time, and also by reducing the accuracy of the predictions made when ML analyzes Big Data.

Incomplete data is usually one of the following:

- *Missing and random:* Data is missing, it follows no particular pattern, but the missing data has minimal impact on the primary dependent variables.
- *Missing but patterned:* There exists a pattern in the missing data that *does*

affect your primary dependent variables.

Typical fixes are as follows:

- Delete all data from files with missing values, and files that contain dummy data. This can only work if the knowledge base is big enough.
- Recover (or correct) the values, either by contacting participants, or by checking other records that contain the same type of data.
- Missing values can be replaced with substitute values in any of the following ways (and more):
  - Educated guessing
  - Calculating an average value from other participants (or data points)
  - Using a midpoint value
  - Using regression substitution to predict the missing value from other values.

Completing Consumer Purchase Data

Oftentimes, consumers who make purchases online will leave non-required fields blank. Marketers will use any or all of the above measures to fill in the missing data. In this case, the marketer has the additional option of contacting the customer for correct info.

Filling In Geospatial Data

Geospatial analysis is a way of applying statistics and other analytic techniques to geographical or spatial data. Geographic Information System (GIS) software is useful for this purpose, as it utilizes geospatial analysis in many ways to manipulate and present all kinds of geographical data. There are a number of ML algorithms for geospatial data. Google Earth (and others) makes good use of geospatial data. Digital navigational maps (in general) are more accurate and interactive than ever before.

Normalizing Temporal Scales Across Data

Some ML algorithms will perform better if the time-series data is consistently distributed. Two techniques that you can use to accomplish this are normalization and standardization.

Normalization refers to restructuring of a database so that all data points are consistent with each other. For instance, if the words *Road* and *Street* are sometimes inputted as *Rd.* and *St.*, to normalize the data, you would need to choose which style you want to use, and change the data accordingly.

To normalize temporal scales across data means to make sure all data has the same numeric time unit, such as a number of seconds, minutes, hours, days, or weeks. Once someone decides which time frame will be used, the important thing is keeping the time units consistent. For instance, if you are measuring time in minutes, a period of an hour and a half will appear as “90.” However, if you are measuring time in hours, then an hour and a half would be displayed as “1.5.”

Standardization refers to another method of rescaling data, by which each piece of data has a mean of 0, and a standard deviation of 1.

Eliminating Seasonality from Data

Seasonality is the existence of variations in time-series data that occur at regular intervals of less than a year, such as weekly, monthly, quarterly, seasonally, or biannually. Seasonality can be caused by any of a number of factors, such as weather, school sessions, and holidays. Seasonality is characterized by repetitive, predictable patterns for a product or service that ebbs and flows over time. But is your downtrend simply an expected post-holiday sales drop, or is it part of something bigger that you need to analyze? Sometimes you may want to analyze a trend without the extraneous “noise” of seasonal data. For whatever reason, you may want your data to be independent of seasonal components. That’s where *seasonal adjustment* comes in handy.

Here is a fairly easy, five-step method for removing seasonality from your data:

1. *Collect data* that goes back at least three full cycle periods.
2. *Compare like time periods.* For example, compare all the months of March, or all Thursdays, or whatever your metric calls for. Then calculate the average of all like time periods.
3. *Normalize your data.* To do this, calculate the average of the averages you obtained in Step 2, and divide each individual average by the result. These are your seasonally adjusted averages.
4. *Divide each original data point by its seasonally adjusted average.* This will provide a new, seasonally adjusted value for each data point.

Normalizing Data Across Different Ranges

All data should be normalized to a range between 0 and 1, to maintain a consistency of values. In other words, marketing professionals must adjust the values of data that was measured on different scales to a single, common scale.

Detecting Anomalies and Outliers

An anomaly, also known as an outlier, is a data point that is at best irrelevant, and may even be harmful to your data set. It is an unusual pattern that does not conform to expected behavior.

Information monitoring services are able to capture and process many hundreds of billion data points each day from web servers, cloud servers, databases, and other digital organizational components. Detection of anomalies allows users to identify deviations from normal levels, even when normal levels change over time.

As we know, ML algorithms possess the ability to learn from data and make predictions. Unchecked anomalies can negatively affect these predictions.

In a very broad sense, there are three main types of outlier data:

- **Point anomalies:** These occur when a single data point is way too far off from other, corresponding data points. In practical use, credit card companies might apply this feature to issue fraud alerts based on the amount of money spent over a certain time period. For another example, an ML algorithm will probably notice if a particular zip code doesn’t match the city or state with which it is normally associated.
- **Contextual anomalies:** These are context-specific abnormalities, commonly found in time-series data. For instance, if someone regularly saves up money all year, in order to go on a summer vacation and splurge, the splurging may seem normal during summer months, but odd at other times of the year.
- **Collective anomalies:** This occurs when people do strange or unexpected things with data, such as sending thousands of emails, or copying data from one machine to another.

Awareness of outliers is very useful in the healthcare sector (for everything from analyzing blood to detecting tumors), as well as in fraud detection and hack detection, to name just some of the many applications of anomaly detection.

Integrating Qualitative and Quantitative Data

Mixed methods research is the collection, integration, and analysis of both qualitative and quantitative data. It is an important step in standardizing a data set.

Weather and Environmental Data

Missing weather data can probably be estimated by way of some clever mathematical algorithm such as Group Method of Data Handling (GMDH). However, here are some simple heuristic approaches to replacing missing weather data that may be appropriate for the rest of us:

- Geographical approach: When weather variables observed at a weather station are incomplete for any reason, it is reasonable to collect data from the nearest weather station that has recorded the weather conditions for the time period missing from your data set. The nearer that nearest weather station is to your weather station, the more accurate your data will be.
- Time-series approach: If the data is missing for only a day or so, you can record the weather for the very next day (or the day before) as the weather of the day that was missed.
- Recollection approach: You can also use your short-term memory to recall yesterday’s weather, and incorporate that thought with today’s recorded weather. For instance, maybe there was a storm yesterday, but today it’s just wet. You might also ask yourself: Did yesterday seem colder, warmer, or about the same temperature as today? Estimate and adjust your missing variables accordingly.

Any of the above approaches may work, although their efficacy depends largely on how precise your weather data needs to be.

8
Applications for Product Innovation

Nobody phrases it this way, but I think that artificial intelligence is almost a humanities discipline. It's really an attempt to understand human intelligence and human cognition.

– Sebastian Thrun, "Reputation in the Age of Artificial Intelligence," contentconnection.PRSA.org, March 16, 2018

Imagine the next generation of Product Innovation:

- No more trend clinics and focus groups.
- No more faith in a bucket of popcorn and a sci-fi movie to know the future.
- Algorithms identify desires and decisions lurking in pieces in the non-conscious.
- Algorithms extract product concepts in the non-conscious and string them together to form a million or so variants.
- Algorithms look for support of the product concepts extracted from the non-conscious in consumers' conscious search patterns and articulation.
- Algorithms extract key category features of the product concept, and predict volumetrics from historical data.
- All aspects of product innovation, pricing, and feature bundling are performed algorithmically.
- Algorithms perform global extraction of new products and trends without ever leaving the computer center.

Rapid advances in AI/ML and Deep Learning will lead product innovation processes to innovate themselves. The tools available right now for that work are bound to evolve in Moore's law timeframes – and faster.

The impact of AI technologies on business is projected to increase labor productivity by up to 40% and enable people to make more efficient use of their time.

– "Artificial Intelligence Poised to Double Annual Economic Growth Rate in 12 Developed Economies and Boost Labor Productivity by up to 40 Percent by 2035, According to New Research by Accenture," https://newsroom.accenture.com/, September 28, 2016

Inputs and Data for Product Innovation

"The only constant is change," the saying goes. The saying is right. Obviously, marketing professionals can benefit from being able to predict these changes in advance, in order to prepare for them and take advantage of them. A trend is an indicator of change. All time-series data sets contain some kind of trend.

Data really is of a few kinds. Non-conscious data refers to data about the non-conscious mind of the consumer. This usually refers to the kinds of music they listen to, the movies they watch, the binge-watched TV shows they consume "mindlessly," the YouTube channels they tune in to late at night, and the Pinterest and Instagram pictures they devour ravenously. Popular culture codes are shaped and fine-tuned through this mega consumption in the non-conscious mind. This popular zeitgeist forms a fundamental and important data source for algorithmic investigation.

Some of this data changes over the course of time, and some of it remains incredibly popular – the Beatles have fascinated generation after generation, and Charlie Chaplin still wows those who discover him.

Western zeitgeist, via pop culture, finds its way to infiltrate and influence minds around the world. New hybrids and variants are created in this confluence of non-conscious cultures.

Then there is information contained in conscious data. This is the data pertaining to conscious human actions: what people speak, write, tweet, blog, search for, post, consider, purchase, advocate for, rant and opine about, and return. These conscious acts contain in them the source and treasure of the next big innovation or desire. In addition, all call center data, focus group data, and consumer survey data contain consumer articulations.

Online search data coupled with real time fluctuations of CPC (cost per click) provide important inputs to innovation.

Demographic data pertaining to emerging, growing, and disappearing segments of the market provide important data for algorithmic exploration.

Global issues, be they environmental, sustainability, or social causes, provide additional information on the cultural zeitgeist that shapes consumer desires.

Desires shaped through aspirations in the non-conscious usually seek validation and support through conscious actions.

Machine Learning and AI analysis for product innovation and R&D requires assembling this necessary data set. The fully assembled data set has a number of elements that cut across brands, geographies, and categories of products.

> *In a survey of over 1,600 marketing professionals, 61%, regardless of company size, pointed to machine learning and AI as their company's most significant data initiative for next year.*
>
> – "Survey Finds Machine Learning and Artificial Intelligence Are Top Business Priorities," https://www.memsql.com/, February 7, 2018

Analytical Tools for Product Innovation

Historically, product innovation required a team of researchers to continually conduct "trend clinics" and perform site visits to "cool places" to determine what consumers desire. Such activities were referred to as "trend hunting" and "cool hunting," and consumed quite a lot of time, energy, and resources. Innovators would then pore over the data that was collected, and brainstorm using human ingenuity to determine what might be interesting innovations. Identified concepts were then presented to focus groups of consumers in standardized concept templates. Consumers were then required to vote on what they thought might be interesting ideas, which would then be taken forward.

Naturally, a great amount of "noise" is injected into this process:

- Trend hunters injected personal biases into what they thought was "cool."
- Reports of trend hunters were not representative of what they observed.
- Generated ideas were not properly worded and fit into the "product concept" template.
- Noise was injected into the consumer survey of the product concept.
- Trend-seeking consumers the trend hunters observed were not the consumers surveyed.

It remains a mystery as to how many wonderful ideas died on the vine as they were swept through this well-intentioned yet archaic process.

A modern Machine Learning and Artificial Intelligence approach to product innovation comprises a nine-step process.

Whether it is an R&D department trying to create new products and services, or an internal service provider like HR trying to see what new services could be offered to employees, these steps for algorithmic exploration prove valuable.

Step 1: Identify Metaphors – The Language of the Non-conscious Mind

Metaphors form the basis of non-conscious thought. The first step in the process is to extract metaphors that pertain to the topic or the focus of innovation. If the category is Soup, then the first step in the process is to algorithmically understand what Soup is a metaphor for.

Algorithmic extraction of metaphors uses semantic frames that structure the bridge between conceptually different domains. So for instance, Soup as Warm Comfort might be a bridge between the domain of Food, and the domain of Feelings or the domain of Home and Parenting.

Such an extraction of semantic frames is possible, and is done by the algorithms of artificial intelligence. These algorithms embody thousands of years of human knowledge, and cultural expressions across the world.

Extracting the global primal metaphors that pertain to the topic at hand becomes the first step of product innovation. Note that in traditional research, this is inferred from thousands of conversations with consumers.

Extractions from consumer conversations are problematic as inferences are drawn from conscious outputs of consumers about the state of their non-conscious. A better approach is to mine the inputs that shape the non-conscious, and draw inferences from them.

Step 2: Separate Dominant, Emergent, Fading, and Past Codes from Metaphors

Metaphors extracted from the non-conscious mind are now validated through data pertaining to the conscious behaviors and actions of consumers. Algorithms look for support in the vast corpus of the conscious for that which was extracted from the non-conscious.

These are nicely grouped based on the level of resonance in both the non-conscious and the conscious.

Dominant metaphors and trends are trends that have already infiltrated our culture. They are hugely resonant in the non-conscious, and find great resonance in the conscious as well.

Emerging metaphors and trends are the trends just around the corner, and they usually have great resonance in the non-conscious, but are not present as much in the conscious corpus. This becomes an easy and clear algorithmic way to evaluate and extract emergent trends. Be they the rise of “naturals” in food ingredients, the rise of Ayurveda in health and beauty, the growth of ETF in mutual funds, these are emergent trends around the corner. They become a vital group to capture.

Fading trends are those that resonate low in the non-conscious, and yet are highly talked about. These are guaranteed to fade, and have short life spans, so building products in line with these metaphors is quite risky. These metaphors allow for rapid and short-term exploitation, however.

Finally there are metaphors and trends that are past. These should be appropriately called the Graveyard of Dead Metaphors. It is important to note that products and services that are built around these metaphors are activities that need a second and third look and must be put to an end.

Step 3: Identify Product Contexts in the Non-conscious Mind

Inputs to the non-conscious mind shape its desires, fuel its fears, and solidify its decisions. Algorithmically identify the particular instances in which the category of product or service appears. Whether it is a fragment of lyric in a song, or a jingle that refers to the category and is lodged firmly in the mind, identify each and every one of these instances where the product or service manifests itself.

Next determine if the context is relevant to the category in question. That is to say that money, if spoken about in the context of economic conditions, is relevant to financial services whereas the context if it is being spoken about as a precondition of romance may be more suitable for dating services. Algorithmic parsing of contexts classifies them into contexts that are category relevant and contexts that are not.

Now further filter these contexts based on their connection to the brand. A brand typically represents itself as a set of functions, a set of feelings, a set of values, and a set of imagery and semiotics.

Once metaphors are chosen, metaphor-based filtering can now be accomplished to extract contexts only where the chosen metaphor is activated as well.

Once these steps are systematically carried out, there exist a corpus of contexts that are ready for algorithmic exploration.

Step 4: Algorithmically Parse Non-conscious Contexts to Extract Concepts

Evaluation of hundreds of thousands of product ideas reveals that product ideas are usually a response to occasions, locations, life pressures, cultural tensions, daily activities, aspirations, disappointments, work pressures, and the like.

There exists a core set of foundational building blocks for innovation that can be algorithmically explored. These building blocks are neither finite, fully fleshed out, nor deterministically defined, but rather are born out of decision and graph analysis of numerous ideas that were successful and an equal number of them that failed.

Algorithmic parsing of contexts feeding the non-conscious mind brings out core creative ideas that can be combined into concepts in an n-gram sense.

Very similar to how neuronal connections are formed in the brain, concepts are chained to form concepts and product ideas.

Step 5: Generate Millions of Product Concept Ideas Based on Combinations

Algorithmic combinations of n-gram concepts result in literally a combinatorial explosion of millions of product ideas and concepts that satisfy a multitude of consumer needs and expectations.

Consumers increasingly expect the products and services they buy to perform more than one task well. One glance at the mobile phone nearby you proves that point; it serves multiple functions, and serves them well. Today, we expect our food to be both good-tasting and good for us, and our cars to be powerful and long-lasting. It is no longer exclusively an “either/or” world we live in.

Engineers can deliver products that feature many capabilities – some to the point almost of overkill, it could be argued. Product designers, and their marketing colleagues, have to determine which of these features to single out for the consumer’s attention. Some of the time that can be obvious. But other times, it can be a vexing dilemma; how to know best which individual or suite of features are going to be the ones that appeal the most to the buyer?

Because AI and ML systems are designed to gather, analyze, prioritize, and synthesize findings on a scale that is Herculean for any human, they can deliver specific findings that product designers and marketers can rely on for guidance when it comes to product feature prioritization. From endless oceans of data in countless databases, AI and ML tools extract the most relevant and meaningful indications of consumer preferences, both conscious and non-conscious in nature.

These systems’ capability to look backward and forward simultaneously, to deduce potential trends before they even fully materialize in the marketplace, to gauge product features’ competitive pros and cons, and even to suggest “counter-programming” measures to help offset a competitor’s product features perceived to be superior by consumers . . . these are all real-world reasons for employing AI and ML in the selection process for product features.

Step 6: Validate and Prioritize Product Concepts Based on Conscious Consumer Data

The millions of n-gram product ideas need validation and prioritization. Algorithmic validation and prioritization occurs for these product concepts through conscious consumer data.

A product concept that is truly innovative will be only lightly referred to in the conscious. If however, no reference is made to it, it might either be entirely undesirable, or just “too far out” for the consumer to even take notice.

Rule-based heuristics, historical innovation success and failure data, and experiential knowledge are coded into algorithms that can then extract, evaluate, and predict the consumer acceptance of a product idea or n-gram. Algorithms may be fine-tuned for a category, for a brand, and for a particular geography based on consumer attitudes.

That is to say, algorithms that look for the next big recipe, the next big ingredient, and the next big health and wellness thing can all be different based on brand and geography.

The next generation of intellectual property are the rule bases and algorithms that perform such evaluations.

Step 7: Create Algorithmic Feature and Bundling Options

Much as with product feature prioritization, AI and ML-based systems can help improve the effectiveness of product bundling. Their ability to assimilate vast amounts of data, from vast arrays of sources, and process it all at blinding speed means that marketers can gain a broader understanding and a larger vision of what particular product bundles are likely to appeal to consumers the most.

These systems also enable marketers to build different virtual models of product bundles and test out assumptions quickly and with more accurate and reliable results.

A product bundle, also known as a “package deal,” describes a marketing innovation where several products are sold as one complete package for a set price.

There are numerous examples of product bundling. For instance, an Office Suite software package typically offers a word processing program, a spreadsheet, a presentation builder, and other features. Priceline offers a discount if you purchase your flight, hotel, and rental car online all at the same time. Restaurants often combine a number of food items into one complete meal (such as the McDonald’s Happy Meal).

Algorithmic bundling is merely a classification of n-grams of product concepts into natural bundles – a product that satisfies A, does B, and addresses C while residing in D. What is typically a laborious process for humans to evaluate and perform is done in milliseconds by a well-trained algorithm.

Step 8: Category Extensions and Adjacency Expansion

While innovation can identify product extensions and service extensions, a very interesting by-product also results – category adjacencies and expansion potential. If for instance the contexts containing the product category in the non-conscious also contained another product category, it clearly indicates a neural binding and association between the two products. As it is in the non-conscious, consumers may be entirely unaware of that binding.

This binding provides a business a wonderful opportunity to diversify and move into an adjacent category. If it might not be strategically prudent to offer that expanded service, it might be useful to partner with another entity to offer the service.

Partnership opportunities, acquisition opportunities, and category expansion ideas naturally arise from algorithmic exploration of contexts in the non-conscious.

Services in conjunction with the product arise when the n-gram of product idea is only partly covered by a product the company offers, and the remaining becomes a service accompanying the product. The Direct-to-Consumer (DTC) approach can always be traced as a service that accompanies the n-gram of product innovation. Indeed, it becomes a powerful way to validate just exactly when a consumer might require that DTC or home delivery as a service.

Step 9: Premiumize and Luxury Extension Identification

Premiumization essentially poses a simple question: What needs to be added to the product or service that will enable it to be viewed as an upgrade and therefore command a better price in the mind of the consumer?

Algorithmic handling of the question is quite simple – figure the n-grams of product innovation, and extract the contexts where the innovation originated. Simply mine the contexts for cues of luxury and premiumization. Feelings, ingredients, and occasions usually provide cues to premiumize.

Product descriptions, or features that connect to the cues of premium, automatically trigger those bindings in the brain, and the consumer subscribes to the notion of premium.

Studies of thousands of products, and core neuroscience, reveal common bindings with premium. Typical cues of premium include, but are not limited to things that have colors of gold, black, red, and pure white. Natural textures and materials evoke premium. Associations with celebrities evoke premium. Time – for instance, whether a cheese is aged or a sauce is freshly made – evokes premium. Bigger, heavier, and stronger evoke premium.

9
Applications for Pricing Dynamics

Artificial Intelligence is not a Man versus Machine saga; it's in fact, Man with Machine synergy.

– Sudipto Ghosh, quoted in "Famous Business Quotes on Success," http://www.bestquotes.io/famous-business-quotes-success/

Imagine Algorithmic Pricing, which has:

- Built in heuristics-based models for pricing, so the user chooses the kind of models to deploy
- Pricing dynamics that leverage all the principles of control system theory
- Pricing control algorithms that dynamically alter price using PID (proportional, integral, and derivative) controllers
- Real-time volume estimators that estimate volume based on systemic, environmental, and dynamic observables
- Daily, hourly, and real-time discounts delivered based on customer and controller outputs
- Psychographic and behavior models that augment and temper strict price volume relationships.

Pricing Dynamics (also known as revenue management) is a pricing strategy where prices change based on real-time customer demand. The cost of a dynamically priced item will vary depending on the time at which the item is sold, the number of unsold items in stock, and/or numerous other considerations.

Pricing is talked about a lot, and yet is poorly understood. While some level of quantitative thinking guides pricing, it is largely a result of category dynamics, consumer confidence, macro and micro economic forces, consumer psychology, competitor actions, and positioning of the product in the mind of the consumer. Many of these variables operate in different time scales, subject to different rules, and do not lend themselves to neat extrapolation or modeling. Human experience and rules play a huge role in determining price. So pricing becomes a wonderful area to explore quantitative machine learning and to capture rules and heuristics humans use from experiences over time.

Price becomes a consumer determinant if no other category-busting metric becomes available. That is to say, that when one competes on price, one is creatively bankrupt. The guidance is to find a metric that captures the core and the essence of the product so price becomes irrelevant.

The rest of this chapter focuses on ways and means of choosing a price that is representative of the forces at work.

> *Every company has big data in its future and every company will eventually be in the data business.*
>
> – Thomas H. Davenport, *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities* (Harvard Business Review Press, 2014), 168

Key Inputs and Data for Machine-based Pricing Analysis

The key inputs to determining price are grouped based on their fluctuations over time. This becomes a smart way to group the variables as the ability of an algorithm to learn and use price changes with the kind of variables used.

Fast time-scale (rapidly changing) data includes:

- Daily weather
- Category average price
- Competitor price
- Consumer propensity to buy
- Consumer loyalty or the lack of it thereof
- Time of day – quantized blocks
- Seasonality
- Time from last paycheck
- Currency fluctuations
- Social media metrics pertaining to product
- Cost per click of category advertising
- Instantaneous demand for product
- Perceived wait time to receive product – supply congestion

- Perceived product shortage
- Stock market and index performance
- Media consumed non-consciously
- Social media output generated consciously

Slow time-scale data includes:

- Consumer confidence
- Economic forces and forecast
- Category and product perception
- Average income of consumer demography
- Sentiment associated with ecosystem of product.

Pricing algorithms would differ based on whether it is a new product being introduced, a variation of an existing product, or just an existing product being dynamically repriced. One assumption is that there is a virtually unlimited number of potential customers, so the size of the population is *not* one of the pricing parameters. Also assumed is that a customer will buy an item as soon as the price is less than or equal to what the customer is willing to pay. Further assumed is that a monopoly situation exists over one's competitors for a particular product or service. Some advantages of dynamic pricing include maximizing profits and pricing flexibility. Among the disadvantages of dynamic pricing are the potential for customer alienation, the advanced technology required to optimally price the items based on real-time demand, and the potential to force competitors to reduce their prices in order to compete.

Dynamic pricing is used on a regular basis across multiple industry sectors – including travel, utilities, insurance, real estate, and entertainment.

While dynamic pricing has numerous advantages, there exist numerous opportunities for it to be inadvertently misused and result in inadvertent consumer discrimination. There needs to be a modicum of caution in the use of algorithmic pricing so it does not discriminate based on age, race, gender, and other parameters that are banned legally.

That moment when you build a bot,

The bot works well,

The bot works too well,

The bot outperforms you in its sleep,

The bot drags and drops your employment contract into the trash,

The bot refuses to write a reference for you,

The bot seduces your wife over email,

The bot helps her find a divorce lawyer,

The bot stops responding to you, and only you.

– *justdickingabout.tumblr.com*

A Control Theoretic Approach to Dynamic Pricing

Pricing may be modeled as a closed loop control system where the output variable is volume of product sold, with a desired volume level to be attained that is set. Now the actual volume is measured in real time, and an error variable which is the difference between the desired volume of product sold and the actual volume of product sold is computed at any time instant. Control action to raise or lower prices is then taken in a manner that is *proportional* to the error, reflective of the rate of change of the error – the *derivate* of the error (and reflective of the cumulative sum of the error from previous time periods) – the *integral* of the error. Thus a P-I-D controller to regulate price is created.

A slightly more sophisticated approach would use a model-based adaptive control approach where the relationship between price and volume is modeled as a dynamical system whose parameters are then estimated and setpoint regulation to the desired volume is achieved through both parameter estimation, and then control law specification.

An optimal control approach does not seek to regulate volume to a set point, but rather tries to maximize a positive definite functional of volume.

A game theoretic approach to pricing seeks a Nash equilibrium through a set of pricing changes with a few assumptions regarding competitor actions in the face of pricing changes.

All control theoretic approaches to dynamic pricing are most useful in online contexts, and pricing through apps and mobile phones be they retail websites or apps like Uber. At the core is a closed loop control system that seeks to regulate or maximize output.

There is a great opportunity, especially in online shopping to let pricing be dynamically controlled through the use of a closed loop control system with certain bound and thresholds in place. Knowledge of the consumer, and therefore their willingness to pay based on items already bought over the course of time, and items consumed in the current session, along with the propensity to respond to a pricing offer can effectively alter the design of the price control system.

Rule-based Heuristics Engine for Price Modifications

Artificial Intelligence engines can take a core price determined through analytical means, and then modify the price in real time using a variety of these heuristics. These pricing concepts can be easily converted into algorithmic embodiments. The following is a quick summary of common pricing concepts. Direct to Consumer models enable using a combination of these pricing concepts in presenting to the consumer.

- **Pricing Floor:** A pricing method in which all costs are recovered, and this forms the floor of the pricing model. The base price is discounted to be at least over the floor.
- **Demand-based "Surge" Price:** Anyone who has taken an Uber knows the meaning of 1.5X – a surge in the base price based on real, or apparent demand. The consumer has no way of validating this claim.
- **Profit Max Price:** "Pharma Bro" tried this approach – simply jack the price up to the point where the consumer is dismayed, but feels there is no other option. ATM fees of banks fall into this category. Eventual regulation and governmental intervention usually follows.
- **Consumer Chooses Price:** A pricing approach by which the consumer is invited to add to the floor and express their appreciation based on service experience. The consumer is invited, but not obligated to pay.
- **Reverse Pop-out Price:** Offer two similarly priced items and one significantly lower, the pop-out. Consumers usually will choose one of the similarly price items.
- **Freemium Price:** A product or service is offered at no cost, then charges premium rates for related products and services, such as advanced features or

increased speed. Most cloud service providers offer such services.

- **Bait and Switch Price:** Wildly discounted items offered for sale to lure consumers in, and moderately priced items offered as the discounted items are "no longer" available.
- **Restaurant Wine Price:** Pricing strategy where the price of an item is simply doubled or quadrupled for convenience. Popcorn in movie theaters also represents this.
- **Competitive Block Price:** This pricing method involves setting a price low enough to discourage competitors from entering the market.
- **Loss Leader Price:** A product that is sold at below cost, in order to encourage the sales of higher priced products. The printer is dirt cheap, but the ink is priced sky high.
- **Odd Price:** This pricing method is one we see all the time: A seller marks the product just below a rounded number. This strategy takes advantage of human psychology to make the price look lower, such as marking an item down from $10 to $9.99. Odd pricing actually does work, no matter how many customers think they aren't being fooled by it. Retail stores and gas station use this extensively.
- **Pay What You Feel Price:** This is an auction strategy. This strategy lets the buyer name the price. This may seem counterintuitive at first, but can be profitable. Parameters can be put in place, such as a minimum reserve price, or a "suggested" price. eBay uses this successfully.
- **Free for the First Six Months:** This method involves setting a price that is either free or ridiculously low to bring customers in, then raising the price later on. The customer has to cancel after the first time period, and if not the regular price will be charged. The customer usually has to share payment information upfront.
- **Ultra Premium Price:** The price is set ultra high in order to evoke favorable perceptions among consumers. This method takes advantage of a consumer's natural tendency to associate high prices with exclusivity and better quality merchandise. This pricing model is a good fit with consumers who really want to keep up with the latest trends. Apple uses this model in its pricing.
- **Contrast Price:** To set the price of one product high, in order to boost the sales of a lower priced product. The price of the lower-cost product will seem even lower in contrast to the higher-priced product.

- **Counter Price:** A form of dynamic pricing where prices are adjusted to correspond with the customers' evolving (or devolving) willingness to pay. This pricing method is often used by online businesses.
- **Demand-based Price:** Setting prices based on customer volume. For instance, an airline might reduce the cost of a flight when takeoff time nears, just to fill the empty seats on the plane.

Machine Learning can then determine which of the pricing schema discussed above were most responsible for volume and associated profits. Note that the schema might differ by category, and some of them may be more suited to an online platform and some may be more suitable for a traditional retail platform.

10
Applications for Promotions and Offers

> *Artificial Intelligence, deep learning, machine learning – whatever you're doing if you don't understand it – learn it. Because otherwise you're going to be a dinosaur within 3 years.*
>
> – Mark Suster, *"Mark Cuban on Why You Need to Study Artificial Intelligence or You'll be a Dinosaur in 3 Years,"* https://bothsidesofthetable.com/, February 7, 2017

Imagine a new world of Promotions:

- Algorithms determine the best context to serve a promotion.
- Algorithms determine the best articulation of a promotion – semiotics and imagery based on the individual consumer.
- Algorithms determine the best context to offer the upgrade, upsell, and convert a customer.
- Algorithms specify the language and neurological codes best intended to convince a consumer.
- Algorithms modify the promotion, pricing, and discounting based on the individual consumer's brand loyalty.
- Switching algorithms determine the best pricing and language to switch a consumer from Brand A to Brand B.
- Algorithms configure in real time the parameters of a promotional offer based on consumer and predicted volume.
- Algorithms dynamically optimize price based on conversion.

Promotions and discounts are one of the most powerful tools to modulate demand. In conjunction with price, they create a compelling call to action.

Machine Learning and Artificial Intelligence can be fully leveraged to explore the power of promotions when adequate data, levers of creative control, and ability to customize and execute them are available. If all three are unavailable, promotions become weak and tired, and never fully result in capturing the promise.

What makes a promotion work might differ as time goes by, but there exist general rules and guidelines to make all of them work. The key learning from analysis of successful and failed promotions reveal that algorithms can play a vital role in ensuring success. The formula for Machine Learning success is usually a variant of the following items:

- Algorithmic determination of the timing of a promotion – path to purchase, location, search intensity – understanding the intensity of intent
- Algorithmic choice of language, metaphor, semiotics, imagery, and of the promotion
- Context scoring in offering a promotion
- Dynamic pricing and calls for action in the offering of a discount
- Clustering and classification of discounts based on success
- Factor analysis of parameters of a discount – amount, percent, expiration, connection of volume, connection to immediacy of payment, connection to advocacy.

> *If we have data, let's look at data. If all we have are opinions, let's go with mine.*
>
> – Jim Barksdale, former Netscape CEO, https://www.goodreads.com/

Timing of a Promotion

The adage "time is everything" is particularly true of promotions. The timing of a promotion and what it is becomes particularly important, not just when a promotion is offered but rather where in the consumer's journey it is offered. The consumer journey typically comprises awareness, information, inquiry, transaction, purchase consideration, advocacy, and enjoyment. In each stage of this journey the consumer seeks different kinds of things.

A promotion merely promotes the consumer efforts in that part of the journey, that is to say it is not reasonable to offer a price discount when a person is merely seeking information. At the time a consumer is seeking information a promotion enables the consumer to consider things to look for and perhaps features that the consumer has not thought about that the product presents. When a consumer is actively considering purchase it is important to offer not just details of the product or promotion but rather to offer the consumer competing information about other products and also the price discount that the seller wants to offer.

In every stage of the consumer journey, these are the seven elements the corresponding promotions must discuss:

1. ***Awareness:*** At this stage, the consumer is barely aware of the brand or product, and is merely becoming aware of the category. Promotions talk of the history of the product, the brand, and generally what other consumers say about it. Algorithms divine the intent of the consumer and offer pithy, attention-grabbing "info teasers."
2. ***Information:*** At this stage the consumer is curious about the category, brand, and product, and is generally browsing for general information, without an agenda or any specific requirement. Promotions talk of the core category-busting metrics of the product. Algorithms divine intent, and offer a single category-busting metric to compare the product to others.
3. ***Inquiry:*** The consumer has specific questions, and these questions contain within themselves the latent needs of the consumer. Promotions compare the product to others in the market and offer general guidance on how best to choose. Algorithms divine latent need of the consumer, and offer key competitive differentiators.
4. ***Purchase consideration:*** At this stage, the consumer is seriously considering purchase and is merely determining the right combination of price, features, and service. Promotions offer price, discounts, warranties, guarantees, bundling, and service offerings. Algorithms create dynamic price and offer discount from real price.
5. ***Transaction:*** The consumer transacts at this point, and can stop a transaction if it appears onerous. Promotions speed up ease of transaction and delivery through loyalty and memberships. Algorithms facilitate "one-click" like ordering based on loyalty, and facilitate familiarity.
6. ***Enjoyment:*** The consumer is getting ready to use the product, and may either be neutral, maximally enjoy the product, or suffer buyer's remorse. Promotions facilitate use of the product, provide service and support, and maximize user experience. Algorithms provide "secret use," "hot buttons," and "tricks and tools." Promotions promise to eliminate the friction in the use of the product – airline lounges and their use eliminate the friction in air travel, and are promotional attributes.
7. ***Advocacy:*** Ecstatic consumers feel energized to talk about the product, and irate consumers feel motivated to slam the experience. Promotions facilitate reviews, offer incentives for advocacy, and turn customers into product

ambassadors. Algorithms facilitate advocacy engines and redeemable incentives.

> *20% of business content will be authored by machines by 2018.*
>
> – Heather Pemberton Levy, "Gartner Predicts Our Digital Future," Gartner, October 6, 2015

Templates of Promotion and Real Time Optimization

Promotional templates grounded in neuroscience typically have the following five-part structure:

1. Opening metaphor
2. Feelings bridging metaphor and connecting to targeted context embedding a promotional concept
3. Targeted promotional concept and call to action
4. Feelings from successful action and connecting to closing metaphor
5. Closing metaphor

Machine Learning algorithms are first used to choose metaphors, and create a small library of metaphors to be served both at the opening and the closing of a promotion. Algorithms similarly create a small library of contexts – occasions, daily activities, life pressures, and so forth that are particularly tuned to the product. These contexts are then to be served with the promotional concept neatly embedded in them. Then a small library of calls to action are created. These calls to action may vary depending on which stage of the path to purchase the promotion is being served. Finally, algorithms serve a metaphor from the library of metaphors to close a promotion.

The use and power of Machine Learning now becomes clear – as data pertaining to the utilization or conversion of a promotion is collected. Now standard techniques (a combination of PCA, clustering, and classification) are applied to the data to determine which opening metaphors, which contexts, promotional concepts, calls to action, and closing metaphors work to maximize conversion. The promotion is optimized in real time or pseudo real time based on the algorithmic learnings in closed loop mode.

> *Heliograf*, The Washington Post's *AI writer, created approximately 850 stories in 2016 during the Rio Olympics and the 2016 presidential election.*

Humanity's saving grace? Editing and analysis polishing came from human editors.

– Capterra Business Intelligence Blog, November 28, 2017

Convert Free to Paying, Upgrade, Upsell

Most services today, be they music streaming, entertainment media streaming, online dating, physical fitness, data storage, doctor home visits, language mastery, grocery delivery, massage, or just plain newspaper subscriptions use simple, brutish models to:

- Offer the service for free for a limited time
- Create interruptions through advertising
- Create hassles for a customer to use the product
- Intimidate a customer to converting into a paying service
- Set up a payment mechanism that is hard to remove
- Hope the customer forgets that monthly payments continue to be made

All of these models crumble eventually owing to either regulatory enforcement, or eventual customer dismay.

The critical question marketers face, to which there have not been many satisfactory answers, is how best to convert, upgrade, and upsell.

Here are observations from the field.

Context matters very much in the world of converting free to paying. That is to say, create a context that non-consciously communicates that you get what you pay for, and that there is pride in ownership, and there is joy in having all of it, and that the consumer deserves to have it all. In this messaging context, introduce the call to convert from free to paying. Machine learning algorithms parse contexts for conversion parameters – be they imagery, music, words, or content – to determine a "conversion index." This conversion index is then used to identify the optimal moment to present a call to action to convert free to paying.

Upgrade and upsell happens in the presence of loss aversion, and "crowd advocacy." Loss aversion is the instinctive cognitive urge that wants to avoid losing "what can be." However, brutal frontal presentation of "what can be" is usually perceived as very sales-like, and is dismissed. It is therefore vital that loss aversion for the upgrade and crowd advocacy for the upgrade are communicated in a subtle and unobtrusive manner.

That begs the question as to how best to communicate the promotional message in a manner that appeals to the non-conscious. Algorithms identify a library of metaphors that have to do with loss aversion for the product and category in question. Libraries of contexts of daily activities, work pressures, or occasions that showcase loss aversion are subsequently identified. Promotional templates are then created that incorporate elements of each of these libraries. Data pertaining to success of upgrades is collected and then simple clustering of the data reveals the means to optimize the promotion for an upgrade.

Language and Neurological Codes

The language and neurological codes of promotion are foundationally different based on age, gender, and culture.

Metaphors seek to create extended conversations while limiting themselves to just a few words or a phrase. They become powerful to activate in the structure of a promotion. Machine Learning algorithms extract metaphors and look for their presence and activation in the conscious communication of humans. Classifying metaphors into *emergent* and *dominant* ones enables further smart use of metaphors. Use dominant metaphors for a promotion of an established product, and consider using emergent metaphors to either revitalize an existing product, or to promote a new product.

The language of promotion contains the semiotics of the underlying metaphor and the imagery conjured by the metaphor. The metaphor clearly dictates the kind of imagery to be used. Neuroscience then dictates the particulars and nuances of the image, and the juxtaposition of words, numbers, and additional concepts with the image.

- The language of promotion geared to men uses directives that utilize paradigms of fight, bleed, and win. This language also uses imagery of games, pursuit, and views from the top.
- The language of promotion geared to women uses collaborative terms that embed emotions as part of its structure. This language uses imagery of people in collaboration and cooperation rather than conflict.
- The language of promotion geared to teens is emotional, using terms that

showcase social belonging. This language also uses a litany of "firsts" – being the first, and being "in the know."

- The language of promotion for 40- to 50-year-olds uses terms that empathize with seeking novelty, seeking newness, and seeking change while also empathizing with seeking fulfillment and peace. Imagery emphasizes restlessness, wandering, and a sense of being lost.
- The language of promotion for 50- to 60-year-olds uses comfort with oneself, and self-acceptance at its core. Gentle humor and the learnings from long journeys form the basis of words and imagery.
- The language of promotion for those aged 60 and above uses positivity, minimality, and clarity. Mischief, joy, and a return to youthfulness form the basis of imagery and words.

Artificial Intelligence algorithms embed the neurological codes into appropriate linguistic context of promotions.

> *44% of executives believe artificial intelligence's most important benefit is "automated communications that provide data that can be used to make decisions."*
>
> – "Outlook on Artificial Intelligence in the Enterprise," *Narrative Science*, 2018

Promotions Driven by Loyalty Card Data

Today, data and customer loyalty go hand in hand to help drive a company's brand strategy. Loyalty-data-driven promotions are based on the assumption that it is less expensive to keep existing customers than it is to find new ones.

The information contained in loyalty data offers a huge competitive advantage, as the most loyal customers can be pinpointed, then targeted appropriately with the right offers and promotions to try to make sure they don't shop somewhere else.

Simplistic use of card data enables targeting an audience with promotional ads for quality products (or product bundles) that specifically suit their individual needs and interests. More sophisticated uses are possible depending on frequency of purchase, basket purchased, and the construction of promotions based on loyalty to the cliched yet useful hierarchy of needs models.

Personality Extraction from Loyalty Data – Expanded Use

Algorithms can systematically map the purchase of individual items to inferences regarding the core personality of consumers; algorithmically map consumer purchase behavior to the common five-factor model; algorithmically tag each consumer to belong primarily to one of the five factors and secondarily to another; and alternatively, cluster groups of consumers to each of the five factors. That is to say, they can tag features of consumer pools to each of the five factors of the model and design promotions based on loyalty card data in one of the following ways, by:

- Promoting new experiences: the curiosity, novelty, and “unusualness” that is uniquely contained in the offer
- Promoting the orderly progression of things: the natural progression, evolution, and disciplined next step in the process contained in the offer
- Promoting positivity and talkability: showcase being the center of attention with positivity and virality by accepting the offer contained in the promotion
- Promoting collaboration and harmony: showcase the collaboration, co-creation, and resulting harmony that arises by accepting the offer contained in the promotion
- Promoting the immediacy and vanishing window for action: showcase loss by not acting immediately upon the offer contained in the promotion.

Algorithmically inferring personality traits from purchase data, and being able to tag consumers to one of the factors in the five-factor model naturally creates the structure of promotions for them.

Charity and the Inverse Hierarchy of Needs from Loyalty Data

Promotions typically target value-seeking customers offering one of the following offers:

- Reduction in price for immediate purchase
- Reduction in price for a bulk buy
- Reduction in price for full payment

- Reduction in price to switch from a competitor
- Reduction in price to place an order
- Reduction in price to use credit

Note that most offers involve a reduction in price. This reduction in price usually confers self-esteem and a bit of self-actualization for the consumer, according to the well-known hierarchy-of-needs theory. Self-actualization occurs as the promotion encourages spontaneity and a bit of creativity; it continues to occur as it confers confidence and accomplishment by “snagging” a good deal, thereby earning the respect of others and giving bragging rights.

It is the general approach of promotions theories that nothing in a promotion promotes the more than the following basic human needs: physiological comforts that have to do with the clothing, food, water, hygiene, sex, and sleep; safety needs that have to do with security of person, work security, security of property, security of health, and resources; love and belonging needs that have to do with friendship, family, relationships, and intimacy. The hierarchy of needs suggests a linear progression of physiological comforts, safety, love, self-esteem, and self-actualization as the ladder of human needs fulfillment. While it is a rational ladder, human behavior suggests the irrational emotional ladder that places needs further up in the hierarchy as being more important.

Charity is a simple connection that connects the top of the hierarchy to the bottom. That is to say, catering to the physiological comforts and safety needs of others satisfies one’s self-actualization and self-realization needs, and thereby facilitates the fulfillment of one’s need for love. This is why, time and again, neurological studies of promotions that involve charitable acts indicate they always work.

Loyalty programs that are built not on providing discounts to customers, but enabling the satisfaction of the physiological comforts and safety needs of others are very powerful. Loyalty card data reveals which of the physiological and safety needs might be useful in structuring charity-based promotions.

> *41% of consumers believe AI will make their lives better.*
>
> – Christopher Dodge, Strategy Analytics, August 31, 2017

Planogram and Store Brand, and Store-Within-a-Store Launch from Loyalty Data

Loyalty card data identifies a few unique aspects of consumers that are underleveraged today. Loyalty data identifies:

- Categories that are consumed with greater rapidity than others
- Categories where diversity of brands increases volume and where it does not
- Categories where introduction of luxury cues and brands increases volume
- Categories where value-seekers crowd, and categories where luxury-seekers abound

Knowing that a lower-income demography seeks a store may be part of the data that supports creating an entire mini store-within-a-store comprised entirely of store brands – devoid of fancy packaging and fancy graphics – thus creating a tremendous value perception. By creating the mini-store within a bigger store, the perceived social stigma of shopping at a “value store” is removed, yet the consumer can gain the value from shopping at a “thrift store.”

Knowing that an affluent demography frequents a store may support similarly creating a mini-store to cater to high-end and luxury items. By just changing layout, textures, lighting, and materials, the perception of luxury is created thereby inviting customers to seek and spend time perusing higher value items. If a category seems to spark consumer interest, it becomes a fertile area for a store to explore with its generic store brands.

Switching Algorithms

How can promotions enable switching a consumer from Brand A to Brand B? Promotions based on price simply become a race to the bottom and rarely win.

Switching happens when the *dominant metaphor* that a competitor has built their brand on, effectively dovetails into an *emergent metaphor* that the product is based on. Thereby, the switch is performed in the non-conscious, where the old dominant metaphor is evolved into a new emergent metaphor. This “sleight of hand” happens at the level of a metaphor, and not by touting conscious attributes of the functional excellence of one product over the other.

Context switching happens in a similar fashion. The life pressures, cultural tensions, occasions, locations, and daily activities that have formed the backdrop of the competitor brand are effectively swapped for the product contexts. Not all of the contexts should be switched, however, as maintaining the tie with the old context might be necessary.

11
Applications for Customer Segmentation

> *As a species, we are very poor at programming. Our brains are built to understand other humans, not computers. We're terrible at forcing our minds into the precise modes of thought needed to interact with a computer, and we consistently make errors when we try. That's why computer science and programming degrees take such time and dedication to acquire: we are literally learning how to speak to an alien mind, of a kind that has not existed on Earth until very recently.*
>
> – Stuart Armstrong, *Smarter Than Us: The Rise of Machine Intelligence*, Machine Intelligence Research Institute, 2014

Imagine the next generation of Segmentation where:

- Loyalty, purchase data, and social listening data for customers is collected
- Principal component analysis or factor analysis determines primary contributors to purchase and defines feature sets
- Segmentation is further fine-tuned based on neurographic information
- Algorithmic segmentation is mapped to primary metaphors; emergent and dominant metaphors in the non-conscious mind
- Algorithms perform real-time segmentation of consumers
- Algorithms perform real-time segmentation of media for advertising and use this data to strategize and rationalize media spend.

Customer segmentation is the marketing process of dividing consumers into targetable demographic groups that have certain aspects in common, such as age, gender, occupation, income, hobbies, or spending habits, to name just a few. Large-scale customer segmentation is made possible by an ML algorithmic capability called *clustering*.

Segmentation has always been a key marketing concept, every bit as important as brand management, corporate identity, distribution, pricing, and research.

These days, marketing is all about customer segmentation. Whether your business is B2B or B2C, segmentation generally allows a more effective distribution of marketing resources to consumers who are more likely to buy the product. Materials sent out using customer segmentation tend to be more valued by the customer, resulting in higher sales overall.

Customer segmentation improves market understanding, allowing marketers to identify new products that their customers might like, as well as to improve existing products based on customer expectations. Segmentation can also help improve a company’s bottom line, as this method can be more cost-effective than mass marketing.

Inputs and Data for Segmentation

Segmentation is a clustering and classification exercise that is data driven. There are many kinds of data that become useful in helping with segmentation. While the data may be obtained, the primary obstacle to successful segmentation is the linking and connection among this multitude of data streams. Aggregate data in each stream seldom matches with aggregate data in other streams. The most powerful way is to match data at an entity level, and then perform reasonable aggregations.

- **Behavioral data:** This is consumer data such as shopping habits, spending habits, loyalty cards, online purchases, credit card data, driving habits, or user status, for example. This data is relatively easy to obtain from large vendors of retail and behavioral data.
- **Contextual data:** This is consumer data pertaining to environmental settings – where they shop, online navigation, locations, occasions, or purchase, time, or proximity to promotion and the like. This data is harder to obtain, and it is quite tough to correlate this with behavioral data. This gets increasingly easier in the world of online marketing, and is quite often the reason for the success of digital marketing. When contexts are identified and at times created, behaviors may be more easily understood and directed.
- **Demographic data:** Typical demographic data comprises elements such as age, ethnicity, income, gender, family size, educational levels completed, and political affiliation.
- **Geographic data:** Geographic range can be as broad as an entire continent, or as pinpointed as a single zip code. Even climate sometimes plays a role when marketers segment customers geographically.
- **Psychographic data**: This is data pertaining to social status, personality traits, political affiliation, values, propensities, and lifestyle choices.

- **Lifestyle and cultural data:** This is data pertaining to choices in lifestyle, relationships, cultural tensions, life pressures, causes supported, sustainability issues, and spiritual and religious choices.
- **Social data:** This is data about websites visited, online habits, online proclivities, social network parameters, and streaming services consumed. This data is becoming increasingly important owing to the convergence of work- and life-related activities all blending together in the digital world.

The output of the segmentation exercise must result in marketing segments that meet the following five criteria:

- The segment must be measurable in some way.
- The segment must be large enough to justify the effort.
- The segment must be somewhat reachable.
- The marketer must have the technological capability to communicate with the segment.
- The segment must be easily differentiated from other segments.

Factors affecting a company's segmentation strategy may include financial resources, competitive activity, and the life cycle of the product, among other things.

Analytical Tools for Segmentation

The five analytical tools for market segmentation outlined next start out deceptively easy, and quite rapidly become dizzyingly complex. There are two primary goals of segmentation:

1. Classifying existing customers so they could be understood differently, and perhaps communicated to differently
2. Identifying new customers for the product or service who have not been addressed.

Step 1: PCA and Clustering Techniques

These are useful to reclassify an existing customer base.

For each of the data sets described earlier, perform a straightforward factor analysis to determine the big influencers of variance, the outcome being purchase behavior. A group of variables will link together to influence variance and identify principal, or form independent combinations of variables that systematically link to influence the outcome.

The factors or principal components specify segments of customers. The mathematics is quite straightforward though the data is quite difficult to get at. Data is required at the individual customer level, and therefore aggregators of vast unique customer data sets are poised to perform this segmentation efficiently.

Alternatively one could perform standard clustering to see how variables clump together in the space of desirable outcomes. Using variations of factor analysis and clustering, data in ONE data set can be arranged and classified.

Note that the data sets present themselves in a transformed space into buckets of clusters. Their interpretation into the original space where the variables live might require interpretation. That is to say 1.5 Female + 0.5 Aged 25–40 is not an easy interpretation for a marketer to understand. Although algorithmically it might make sense, it would require further interpretation to know what actions might need to be taken. Marketing data rarely falls neatly into clusters and algorithmic outputs without further human intervention to interpret and fine-tune the projected clusters, especially the actions to take.

This exercise is repeated for each of the data clusters, with a separate segment arising from each one of them. Rarely are data sets integrated to arrive at a single data stream across all data sets. That is to say, the best kind of data is where for a single customer, there is data across all data sets. Integrating retail behavior data and media consumption data across a single customer can create a unique analytical advantage.

Note that each data set, be it behavioral or demographic, might come up with an independent set of clusters. These will then have to be further integrated to result in a meaningful and finite set of segments that enable actions.

Step 2: Metaphor-based Segmentation

This method is useful to identify new customer segments.

Metaphors become a powerful instrument to understand and segment customers. Segmentation exercises based on metaphors serve to identify newer segments not based on past purchase habits but on the proclivities to purchase if presented to properly. Algorithmic extraction and prioritization of metaphors make such segmentation possible.

First, extract metaphors pertaining to a category or product. These metaphors are algorithmically extracted by understanding the semantic frames pertaining to a metaphor. Next look for the resonance of the metaphor both in the non-conscious, through the inputs to the non-conscious mind, and for the resonance in the conscious, as outputs of the conscious mind. Arrange the metaphors by rank order to find emergent and dominant metaphors.

Note that each metaphor represents a cluster of non-conscious thought. So Taste as Pleasure or Taste as Sophistication or Taste as Mischief represent fundamentally different customer segments.

Now that the metaphors have been identified, for each chosen metaphor, identify consumed media that contains the metaphor and the category – that is to say, Taste is Pleasure in conjunction with either Ice Cream or Soup, depending on the category of interest. Demography that consumes the media is the target demography for the product. Identify conscious outputs that contain the chosen metaphor and the category. The preferences in conscious outputs represent the preferences of the desired segment.

Tailor communication and media to target the desired segment.

Step 3: Algorithmic Facet-based Segmentation

Another step is to algorithmically extract contexts that contain the product and category in media consumed by the non-conscious. If it is a new product, utilize key descriptors of the product to perform the extraction. Perform a similar extraction of contexts in outputs of consumers be they social media musings, blog posts, Amazon reviews, or call-center transcripts.

Now look through this corpus to identify a variety of facets such occasions, times of day, daily activities, sports, fashion, and fitness that are present in the contexts containing the product and category.

Look for n-grams from different facets that are contained in these contexts. These n-grams contain all necessary and sufficient information pertaining to consumer segments that are potential targets for a new product. Non-conscious media consumed reveal the target demography.

As an example, if Breakfast and Soup are contained in a set of contexts, and the same context contains Life Pressure of Single Parenting and Work Pressure of Late to Work, one might hypothesize that one demography to target might be Overworked Single Parents.

Step 4: Segment Fusion Based on Plurality of Approaches

Akin to sensor fusion when sensors present different albeit incomplete views of reality, segmentation using quantitative, metaphor-based, and facet-based approaches yield slightly different views of reality.

A way to fuse these approaches is based on the primary business driver. The search for new segments is rarely well accomplished by using the same approach on the same data. So the use of metaphors and the use of algorithmic facets might yield results. Seeking to revitalize an existing customer base by understanding these facets differently might require a more quantitative split of the “well understood pie,” and so the quantitative approach might suffice.

A newer wave of segmentation is to use a hybrid approach of blending all of the aforementioned methods.

Step 5: Segment-specific Offerings

The modern approach is to offer product features, product bundling options, service options, and even business models based on identified segments.

Segment-specific consumer experiences are the next “big thing.” Be it experiencing the product, or experiencing the world using the product, the offering of personalized and customized consumer experience that accompanies the product changes the way marketers approach the consumer.

Among the numerous experiences that are possible with the product, the intelligent use of metaphors and the algorithmic facets enable a level of hyper-personalization previously not thought possible.

The extraction of semiotics from the metaphor can influence even the language of communication with the segment be it through a call center, a chat bot, or a direct personal extraction.

12
Applications for Brand Development, Tracking, and Naming

> *In case you are sitting here pondering this question thinking that AI will never eliminate human intelligence because humans still have to program and train them, that isn't entirely true. Right now, there are of course still researchers, programmers, and engineers who train robots and rudimentary AI systems. However, more and more code – much of it in relation to AI – is actually being written by AI programs already. Programmers today no longer have to write long complex codes for AI telling the robot to do this or that. They simply have to write code that tells a program to write code telling the AI to do this or that.*
>
> – Trevor English, "Will Artificial Intelligence Spell the End for Human Intelligence?," interestingengineering.com, March 31, 2017

Imagine developing and tracking brands with:

- Brand understanding using Big 5 tests
- Brand tracking using Big 5 tests
- Activating brand identify in social conversations
- Resonance of brand identity in the non-conscious media consumed
- Celebrities whose work, personality, or characteristics connect to brand identity
- Tracking of reputation of celebrities, and social media influencers
- Connection of product names to metaphors and brand identity
- Extraction of contexts from the non-conscious connected to a brand
- Clustering of brands based on function, feelings, and/or values
- Factor analysis of brands in a category

Brand Personality

Brand Personality development and tracking has a fundamental problem – the typical dimensions of Brand Personality development do not correlate or correspond to the commonly used personality understanding tests such as the Big 5 – or the 5-factor tests. This usually leads to segmentation issues, and a lack of understanding of which attributes of brands might naturally correspond to attributes of a consumer's personality. Development of creative messaging is further complicated as a result of these disconnects.

Knowing the purchasers of brands, and the personality of the brand they purchase, it is possible to extrapolate from loyalty card data the personality of purchasers – individually, in aggregate, by geography, by brand, by category, and by retailer. This algorithmic extraction provides the possibility of tailored messaging, as well as brand development and promotional offerings.

Traditional brand personality development comprises scales such as Sincerity, Excitement, Competence, Sophistication, and Ruggedness. These are further broken down into facets. Brands are then tracked according to these facets and scored over periods of time.

Brand tracking is quite rudimentary today – social mentions of brand are tracked along with a crude measure of sentiment associated in the mention. These measures are largely unsophisticated, and do not lend themselves to targeted action.

These measures of the personality of a brand do not easily reconcile with the typical dimensions of human personality as measured by the Big 5 tests. The Big 5 personality traits have robust psychometric validation, are quite universal in their applicability, and lend themselves to statistically validated measurement. The scales are well understood, and measurement is along this scale.

The Big 5 personality traits seem to be quite universal and cut across cultures and geographies. There is emerging consensus that these traits may have biological origins and may have shaped social evolution. Interesting studies of human twins suggest very statistically significant inheritance of these five traits that raises the interesting possibility of biological and genetic predispositions for these traits. Longitudinal studies of these traits suggest their relative stability in adulthood with no statistical changes over five-year periods despite occurrence of life events. Movement happens along all of the scales past 60 and settles out. There is increasing evidence that social groups are clusters of people with similar trait scale measures.

It is important to note that each of the five personality factors represents a range between two extremes. The brand personality is a scale between the two extremes along each of the five factors.

In the real world, most people and brands lie somewhere in between the two extreme ends of each dimension. The very same way social networks become networks of people with similar personalities, brands when described as people end up with people with similar personalities.

Machine learning enables drawing the following inferences:

- Knowing brands bought, infer personalities of purchasers
- Knowing personalities of purchasers, infer brands they may be interested in
- Craft the personality of a brand, and be able to infer and identify potential purchasers

These five Brand Personality archetypes may be described as follows.

Brand Personality Type 1: New Experiences and Openness

Brands on one end of the scale showcase products and services that have creativity, cater to curiosity, and have features that bring out a consumer's spirit of adventure and exploration and a desire for change. Product announcements for these brands tend to be looked for with great excitement, and product features are constructively disruptive, and break rules and traditions of product design. Brands on the other end of the scale showcase products that honor tradition; preserve heritage, heirlooms, and old ways of doing things; and resist change.

Brand personality scales range from Creative, Explorative, and Adventurous to Traditional/Tried and Trusted.

Brand benefit scales range from Abstract and Ethereal to Comfort and Knowing.

Brands that score on one end of the openness continuum are typically:

- Very creative – in design and options
- Open to trying new things
- Create products that address unknown challenges
- Convey benefits that are more abstract than tangible

Brands that score on the other end of the openness continuum are typically:

- Traditional and conservative

- Preserve old ways of doing things
- Resist change for change's sake
- Solid and grounded
- Convey benefits that are known, endure, and tangible

Brand Personality Type 2: Orderly Progression and Conscientiousness

Brands on one end of the scale showcase products and services that have thoughtfulness, are evolutionary in nature, involve rules, measurement of progress, feature modules, and platforms with a sense of progression through the programs with specific goal directed behaviors. Brands on the other end of the scale showcase products that allow for spontaneity, features that allow for messes and spills, and generally accept that humans will not do what they are supposed to do and ably compensate for these.

Brand scales range from Discipline and Progression to Spontaneous and Unplanned.

Benefit scales range from Scheduled and Detailed to Unscheduled and Unstructured.

Brands that score on one end of the conscientiousness continuum typically:

- Have products and services that require time to prepare and use
- Have a process and rules to follow
- Pay attention to little details
- Involve organization, structure, time, and schedule

Brands that score on the other end of the conscientiousness continuum typically:

- Have products that facilitate spontaneity
- Allow for mess, spills, and accommodate them
- Utilize disposable elements that do not require care or re-use
- Relax time–based use and scheduling
- Autocomplete, autocorrect, and accommodate user errors

Brand Personality Type 3: Positivity, Talkability, and

Extraversion

Brands at one end of the scale showcase products and services characterized by excitability, high energy, sociability, talkativeness, assertiveness, and high amounts of emotional expressiveness. Brands enable commanding attention, and being the center of attention. Brands at the other end of the scale showcase products that are subtle, nuanced, understated yet elegant, and represent thoughtfulness and depth.

Brand scales range from Outward Excitement and Sophistication to Inward Quietness, and Subtle and Deep.

Benefit scales range from Virality and Talkability to Nuanced and Understated.

Brands that score on one end of the Extraversion continuum typically:

- Create products and services that become the center of attention and are likely to start conversations
- Create and convert new customers
- Focus on creating a large fan base
- Make bold claims regarding products and services

Brands that score on the other end of the Extraversion continuum typically:

- Stand alone in being understated and nuanced
- Exude quiet and confident energy
- Say little and say it with subtlety
- Eschew small talk, generic and trite statements
- Deliberate about design and functionality
- Products and services do not call attention to user

Brand Personality Type 4: Collaboration, Harmony, and Agreeableness

Brands at one end of the scale showcase products and services characterized by trust, social justice, values, honesty, kindness, and fairness. Collaboration, and cooperation and betterment of all are emphasized. Brands at the other end of the scale showcase aggressive competition, and a willingness to be ethically flexible to win and get the best deal for the consumer.

Brand scales range from Kindness, Empathy, and Social Justice to Competitive and Aggressive.

Benefit scales range from Sustainability and Fairness to Competitive and Ethically Flexible.

Brands that score on one end of the Agreeableness continuum typically:

- Create products and services based on sustainable business practices
- Use labor practices that are based on social fairness
- Create products that carry ethical and best-practice certifications
- Focus on corporate giving back and improving society

Brands that score at the other end of the Agreeableness continuum typically:

- Focus on beating the competition – through low prices, value, speed, or sheer aggression
- Are ethically flexible in the pursuit of beating competition
- Extract value from suppliers and producers aggressively
- Are less focused on sustainability or fair business practices
- Unbound by rules
- Win at all costs

Brand Personality Type 5: Emotional Volatility and Neuroticism

Brands at one end of the scale showcase products and services characterized by emotional volatility—emotional highs and emotional lows, that is to say there is an emotional rollercoaster associated with products and services of this brand. Brands on the other end of the scale showcase products that have emotional stability, are grounded, resilient, and provide the rituals and constants in life.

Brand scales range from Emotional Volatility and Neurotic to Grounded and Stable.

Benefit scales range from Emotional Highs and Lows, to Stable and Sane.

Brands that score on one end of the Emotional Volatility continuum typically:

- Create products and services for competitive and stressful activities

- Associate win/lose with products and services
- Are geared for a stressful lifestyle
- Embrace emotional highs and lows rather than eschew them

Brands that score on the other end of the Emotional Volatility continuum typically:

- Create products and services that provide emotional stability
- Create products that through mundane ritual create groundedness
- Eliminate worry and anxiety through the familiar and simple
- Focus on creating a sense of ease and relaxation

> *Thirty-eight percent of consumers said they believe AI is going to improve customer service.*
>
> – "What Consumers Really Think About AI," Pega, https://www1.pega.com/system/files/resources/2017-11/what-consumers-really-think-of-ai-infographic.pdf

Machine-based Brand Tracking and Correlation to Performance

Identifying a brand personality along the dimensions of the five factors creates a unique psychometric scale that is well understood among humans. The extremes of each of the scales are characterized by a set of attributes that lend themselves to linguistic assessment described by the following steps.

Step 1: For each of the ends of the scale of each of the personality traits, extract a set of metaphors that represent best the end of the scale.

Step 2: For each of the ends of the scale of each of the personality traits, extract the linguistic representation that adequately captures the end of the scale.

Step 3: Algorithmically comb through social media mentions of the brand and identify activation of endpoint metaphors, and distance from them for each trait.

Step 4: Algorithmically comb through social media mentions of the brand, and identify distance from linguistic representations of endpoints for each trait.

Step 5: Perform a quantitative blending of the measures to get at the score of the brand along each trait.

Each brand is now scored along the Big 5 traits algorithmically, globally. Performance of brands is best understood by performing a factor analysis of the scores along each of the personality traits to the performance of a brand. This analysis provides core data of how brand leaders in a category have positioned themselves. Brand creative strategy and messaging strategy can then be altered based on the analysis.

Continuous tracking of the brand along these dimensions also reveals the impact of creative messaging and media performance.

Machine-based Brand Leadership Assessment

Brand leadership is easily understood and tracked based on the studies of correlations of human leadership with the Big 5 personality traits. The correlates of brand leadership, based on correlates in human leadership in the Big 5 personality traits, are as follows:

- **Extraversion – highest correlation:** Extraversion is a predictor of emergence, not necessarily sustained performance. Sociability and talkability in particular on one end of the scale are highly correlated.
- **Conscientiousness – second highest correlation:** Being organized and hard working with rules and procedures to follow, conscientiousness again is a predictor of emergence, not necessarily sustained performance.
- **Openness – third highest correlation:** At times, however, openness is valued as highly as extraversion.
- **Neuroticism:** This trait is the least correlated with leadership.
- **Agreeableness:** This trait is not well correlated to emergence, but highly correlated with sustained performance.

Brand leadership measurement is translated to algorithmically tracking these attributes of human leadership along these dimensions.

Loyalty card data provides further insights into how brand leadership, and therefore performance might differentiate itself by category. So while broad generalizations may be made, loyalty card data provides an indication of how microcosms of demography might view the brand traits differently, and it therefore explains brand performance.

The next generation of brand-to-consumer correlates would also include as inputs the Big 5 personality tests of panelists. Knowing the personality traits of consumers, and the personality traits of brands they favor allows for more personalized messaging and deeper construction of brand identity.

Machine-based Brand Celebrity Spokesperson Selection

A successful celebrity endorsement can help improve brand and product awareness, point-of-sale recall, purchase motivation, and it can help increase customer loyalty. Celebrity endorsements are increasingly “transnational” – top Hollywood stars, who would not necessarily choose to appear in US commercials, are highly popular spokespeople for brands in markets like Japan.

Algorithmic mapping of a celebrity along the 5-factor model can be extracted from social media postings involving the celebrity. That is to say, a celebrity is a value along the five dimensions of the 5-factor model.

There then exist two ways to utilize a celebrity:

- Exact Match – when the celebrity accurately maps to each of the scales of the five factors very close to what the brand has mapped itself, this is an exact match and therefore creates resonance. The celebrity amplifies the attributes of the brand, and becomes an accurate brand ambassador.
- Compensatory Match – when the celebrity facilitates movement from one end of the personality trait to another. It is important therefore to choose a celebrity who represents the migration along the scale, but is not so far ahead that there is a credibility gap.

Machine-based Mergers and Acquisitions Portfolio Creation

Akin to choice of celebrity, M&A has two possible foci – either created depth in a portfolio of brands, or created breadth in a portfolio of brands.

This strategic objective is algorithmically captured in the following manner:

- Exact match of the personality of brand under consideration to other brands in the portfolio creates depth to the portfolio, and assists with retention of loyal users.

- Mismatch of the personality of brand under consideration to other brands in the portfolio creates breadth to the portfolio, and assists with capture of new users.

The success of a merger or acquisition depends a lot on brand considerations. Do we keep one name and lose the other? Combine the two names together? Discard both names and dream up a new one? With M&A, there are so many detailed and important decisions to be made regarding the brand portfolio. These questions are typically resolved in the following manner – portfolio deepening requires removing the name of the new brand, and portfolio breadth expansion requires retaining the name of the new brand.

Machine-based Product Name Creation

There are many AI-assisted creative name generators that can suggest names for brands, products, services, digital assets such as websites and blogs, and companies, based on inputted criteria. Some of these name generators are quite sophisticated, and can really help marketing professionals think “outside the box.”

Brand and product names generally fall into one of the following categories of algorithmic expressions (from A to T – acronyms to translations) listed below. When you consider the enormous range of naming possibilities that these categories – individually and collectively – represent, the value of applying AI- and ML-powered algorithms to sort through, analyze, and synthesize those possibilities becomes self-evident, and the following steps apply.

Step 1: Identify for the brand/product the desired attributes along the Big 5 Factor.

Step 2: Where the brand/product falls toward the ends of the scale, and only for those traits, identify the endpoint metaphors and endpoint linguistics.

Step 3: Based on where the brand lands along the five factors, identify other brands that have a similar profile, and use their names.

Step 4: Treat every word and concept that comes from Step 2, and run it through the algorithms described below to generate a set of names.

Step 5: Using loyalty card data, identify all names of brands valued by the target demography and identify the algorithmic output most suited for the target demography.

The following are well-documented methods easily found in Wikipedia; they lend themselves to algorithmic expression using a combination of morphemes, phonemes, and syntax. Note that not all of them can be algorithmically enabled. The ones that can be algorithmically enabled are noted.

- Acronyms – IBM, KFC – can be algorithmically enabled
- Amalgams – Nabisco – can be algorithmically enabled
- Alliteration – Best Buy – can be algorithmically enabled
- Appropriation – Caterpillar
- Blend – Pinterest – can be algorithmically enabled
- Clever Statement – The Boring Company
- Clipping – FedEx
- Descriptive – General Electric
- Eponyms – Trump Tower – can be algorithmically enabled
- Evocative – London Fog
- Founder – Hewlett-Packard, Ferrari
- Geography – California Pizza Kitchen – can be algorithmically enabled
- Homophone – Krispy Kreme
- Ingredients – Spice Market – can be algorithmically enabled
- Merged – ExxonMobil – can be algorithmically enabled
- Mimetics – Google
- Morphological Borrowing – Nikon
- Nicknames – Kinkos
- Onomatopoeia – Ring – can be algorithmically enabled
- Omission – Razr – can be algorithmically enabled
- Personification – Nike – can be algorithmically enabled
- Poetics – USA Today

- Portmanteau – Travelocity – can be algorithmically enabled
- Prefix – iPod – can be algorithmically enabled
- Reduplication – Spic and Span
- Rhyme – YouTube – can be algorithmically enabled
- Removal – Acura – can be algorithmically enabled
- Replacement – Vimeo
- Suffix – Printify
- Synechdoche – Staples
- Theronym – Mustang
- Translation – Volvo

13
Applications for Creative Storytelling and Advertising

> *Despite how it's portrayed in books and movies, artificial intelligence is not a synthetic brain floating in a case of blue liquid somewhere. It is an algorithm – a mathematical equation that tells a computer what functions to perform . . . In the world of AI, the Holy Grail is to discover the single algorithm that will allow machines to understand the world – the digital equivalent of the Standard Model that lets physicists explain the operations of the universe.*
>
> – Jeff Goodell, "Inside the Artificial Intelligence Revolution: A Special Report, Pt. 1," *Rolling Stone*, February 28, 2016

Imagine creative storytelling that has its core:

- Metaphor extraction – semantic frames drive
- Algorithmic extraction of semiotics and imagery
- Algorithmic extraction of creative facets such as locations, occasions, life pressures, cultural tensions, etc.
- Algorithmic map of brand identity, metaphors, and extracted creative facets into creative story lines
- Algorithmic map into creative templates for online delivery
- Copy testing morphs into algorithmic context scoring
- Algorithmic extraction of musical elements to match sentiment
- Algorithmic extraction of musical contexts to match copy context
- Imagery map to online digital expressions without sound
- Real-time PCA and factor analysis to determine principal influencers of template online advertising.

It can be argued that marketing is not so much about the product (or service) that is produced, as competitors often produce similar products. Marketing is more about the *story* that is told – about the brand, about the product, about the consumers who use the product. We often think in terms of stories, whether or not we are aware of it.

Storytelling creates an emotional connection between a consumer and a product or brand. Ideally, each story increases the emotional attachment between brand and consumer. Neuroscience teaches that looking for and identifying *context* is a constant, critical path that our brains take, in order to make sense out of our world. Storytelling is a way of establishing context.

Compression of Time – The Real Budget Savings

Real budget savings are achieved by reducing the biggest expense item – media-buying expense. If the same message can be conveyed more effectively in a shorter amount of time, it benefits both the consumer and the marketer.

With complexity of products and messaging that becomes a problem: How to convey a longer message in a shorter amount of time?

Various studies of advertising effectiveness have found a very interesting correlation – the entire effectiveness of an ad seems to be highly correlated with the effectiveness of approximately the first five seconds. How to be effective in just five seconds where neither the story, nor the characters or the brand, have been yet introduced? This becomes the second dilemma of the marketer.

Another challenge for today's storytellers is the rapidly waning human attention span. "Insta"gram, capturing the "instant," and "Snap"chat capturing the message in a snap, as well as the ubiquitous tweet, have made long-winded discussions a thing of the past. According to some studies, human attention span has decreased, on average, about 33% over the past 15 years. This is one reason why Snapchat stories are limited to 10 seconds in length, and infographics are in fashion. If you fail to capture the attention of the people in your digital audience within the first three or four seconds, you are likely to lose them. Consumers are also loath to watch anything that exceeds seven seconds in its entirety – so telling a story in seven seconds or less is a challenge.

These sets of challenges are mitigated using algorithms to extract metaphors that pertain to the message and the brand, using algorithms to identify the contextual elements in the consumer's mind already connected to the message and brand, weaving a story or message that conforms to neuroscience templates that are guaranteed to connect to the consumer, and using the power of the brain to tell the story. Impressionists painted dots, and utilized the power of the brain to connect to the dots to have water lilies emerge in the gardens of Giverny in the canvasses of the mind.

The following process is helpful in constructing the next generation of time-compressed advertising and storytelling for a given topic, and a given brand:

Step 1: Extract metaphors that are culturally, historically, and archetypically connected to the topic. This is algorithmically accomplished using algorithmic metaphor extraction engines.

Step 2: Look for the support of these metaphors in the media consumed by the non-conscious mind, and support for them in the outputs of the conscious mind. This is algorithmically accomplished using engines of language understanding.

Step 3: Identify and isolate the metaphors that are *emergent* – scoring high in the non-conscious media consumed, and scoring low in conscious outputs. Identify and isolate metaphors that are *dominant* – scoring high in the non-conscious media consumed, and scoring low in the conscious outputs. This is algorithmically accomplished using language understanding and statistical ranking.

Step 4: For each of the identified emergent and dominant metaphors, find which of them are already explicitly connected to the brand in the conscious outputs of the consumer. This is algorithmically accomplished by contextual understanding.

Step 5: Determine if any emergent or dominant metaphors already connected to the brand are to be retained.

Step 6: Based on the topic of messaging, determine which of the emergent or dominant metaphors are to be activated in the messaging. Add a brand filter if need be to match the semiotics of the chosen metaphor with the key characteristics of the brand. This further refines the permissible list of metaphors connected explicitly with the brand and its personality.

Step 7: For the chosen set of metaphors to be activated in a campaign on the topic by the brand, identify the semiotics, imagery, and core essential linguistic profile of the metaphor.

Step 8: Identify contextual elements connected with the topic in the media consumed by the non-conscious mind. These contextual elements may be daily activities, occasions, locations, and times of day, or cultural tensions and the like. These are algorithmically extracted.

Step 9: Rank order the extracted contextual elements based on the support for the same elements in the conscious musings of consumers. Algorithms provide this statistical parsing.

Step 10: Identify contextual elements that have deep resonance in the non-conscious mind, but have not become clichéd and therefore subject to repetition blindness. Algorithms perform this identification of deep resonance.

Step 11: Create a storyline that embodies the chosen metaphors and utilizes the contextual elements to tell the story.

Step 12: Based on identified contextual elements, and topic, pick the piece of music that lyrically connects best. This is performed algorithmically.

> *Big Data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.*
>
> – Dan Ariely, Facebook, January 6, 2013

While there exist many paradigms and templates to create, the following have been found to be consistently effective in communicating the message. Note that the template does not specify what the message should be, but merely strives to showcase the architecture of the package that facilitates easy absorption by the brain of the consumer.

Template for Constructing a 30-Second Ad

1. Opening metaphor using semiotics and imagery to connect to the non-conscious mind in less than 5 seconds
2. Emotional components of story and message using identified contextual elements in the first 10 seconds
3. Functional components of message using identified contextual elements in the first 20 seconds
4. Emotional call for action in story using identified contextual elements in the first 25 seconds
5. Closing metaphor with the brand in the last 5 seconds

Template for Constructing a 15-Second Ad

1. Opening metaphor using semiotics and imagery to connect to the non-conscious mind in less than 5 seconds
2. Emotional and functional components of story and message using identified contextual elements in the first 10 seconds
3. Closing metaphor with call for action in the last 5 seconds

Template for Constructing an 8-Second Ad

1. Opening metaphor using semiotics and imagery to connect to the non-conscious mind in less than 5 seconds
2. Emotional call for action in the last 3 seconds

Template for Constructing a 5-Second Ad

1. Opening metaphor and call for action in 5 seconds

Template and Components for Constructing a Print Ad

1. Metaphor that uses semiotics and imagery to connect to the non-conscious
2. Emotional and functional components of message using identified contextual elements
3. Call for action

Template and Components for Constructing an Internet Banner Ad

1. Metaphor that uses semiotics and imagery to connect to the non-conscious
2. Functional components of message using identified contextual elements
3. Call for action
4. Algorithmically switch semiotics choices, imagery choices, call for action choices, and contextual elements based on click-through and conversion rates in real time

Template and Components for Retail POS

1. Metaphor that uses semiotics and imagery to connect to the non-conscious
2. Emotional components of message using identified contextual elements

3. Call for action
4. Algorithmically switch semiotics choices, imagery choices, call for action choices, and contextual elements based on sales in real time for digital signage and in pseudo real time if not

Weighing the Worth of Programmatic Buying

The importance and power of metaphors is clear through their consistent use in both the opening and closing of the ad.

There is considerable, and at times questionable, investment in programmatic buying. It is argued that programmatic purchase works best when there is alignment of context and only when the metaphor pertaining to either the category or the brand is activated. That is to say that the closer the context is to the metaphor, the more receptive is the consumer to an activation and therefore represents the “best purchase.”

Note that this means of purchase allows for specific exchange of information between buyers and sellers. It is argued that without this exchange of information, the transaction is flawed, and the marketplace is inefficient.

Programmatic Advertising Purchase Logic

If

Metaphor is activated

AND

Key functional contextual elements are present

AND

Key emotional contextual elements are present

THEN

Purchase spot paying maximum allowed price

ELSE

Do not purchase spot

OR

Reduce maximum allowed price

Template and Components for Meme Construction

A meme is a cultural idea, behavior, symbol, or style that is virally shared by way of a series of imitative variations. Although some memes are meant to ridicule stupid human behavior, most memes are meant to be amusing. Memes have become a global social phenomenon, the best ones achieving viral status by way of social media.

The key observation is that the metaphor is the meme.

- Emotional depiction of the source frame of the metaphor – using a reduced semiotic set, and associated imagery
- Action depicting the metaphor
- Emotional depiction of the destination frame of the metaphor – using a reduced semiotic set, and associated imagery

Use of metaphors has the potential to dramatically impact effectiveness while at the same time cut down on the time needed to deliver the message. Algorithmic extraction, classification, and activation of metaphors and their associated semantic frames and structures become the next frontier in the use of Machine Learning and Artificial Intelligence in creative storytelling.

> *Could you lose your job to AI? According to PwC, maybe. By the 2030s, they predict that around 38% of all U.S. jobs could be replaced by AI and automation.*
>
> – Richard Berriman and John Hawksworth, "Will Robots Steal Our Jobs?," UK Economic Outlook, March 2017, https://www.pwc.co.uk/economic-services/ukeo/pwcukeo-section-4-automation-march-2017-v2.pdf

Neuroscience Rule-based Expert Systems for Copy Testing

Neuroscience research has shown that the vast majority – 95% – of purchasing decisions is driven by the non-conscious mind.

The sheer power of that metric underscores how critical the non-conscious mind is in our daily lives. Understanding some of the core drivers within the non-conscious mind provides the keys to creating effective innovations in marketing and new product development.

While there are a number of elements that command the attention of the non-conscious mind, a few deserve special attention in the world of diminishing consumer attention. These lend themselves to rule-based expert systems to algorithmically score and transform the world of copy testing.

- **Motion score:** Motion commands attention, and mesmerizes the human brain. Motion in the early parts of an ad, or on an internet banner, or in the retail POS, commands consumer attention.
- **Novelty score:** Unexpected newness commands the attention of the brain. A surprise, such as the opening of a curtain or the unveiling of something in an advertisement or in a user-induced action, is pleasurable for the brain.
- **Error score:** Simple errors that are easily corrected are pleasurable for the human brain. So any TV or internet advertising that has a small error that a user's brain can correct very fast without too much effort is found to be wonderful.
- **Ambiguity score:** Things that are partly covered and partly revealed are pleasurable for the brain.
- **Implicit humanity score:** Pure functional descriptions of the product or process in which the science or technology of a product are discussed with no mention of humanity are discarded by the brain. The presence of humans is what commands attention.
- **No cortisol beginnings score:** Advertising that begins by stating problems and thereby creating stress and a case for change do not typically win the non-conscious mind's positive retention.
- **Voice-over score:** Female brains prefer being spoken to by female voices.
- **Sound score:** While all sound intrigues the brain, voices of children, whispers, moans, and groans command attention as they implicitly contain emotion.
- **Music score:** Music priming happens in the brain between the ages of 15 and 22. The score determines the appropriateness of the age of the music with the age of targeted demography, especially in the early seconds of an ad.
- **Lyric semiotic score:** Scores are achieved if the semiotics of the lyrics of the chosen music connects with the semiotics of the metaphor and the creative thrust of the ad.
- **Optical illusions:** This is a continuation of the brain's fascination with novelty or unexpected newness. Optical illusions in which what you perceive

is more than or diametrically opposed to what you see command the brain's attention and are particularly useful in digital advertising and retail POS.

- **Slow motion:** The brain is obsessed with slow motion. Sheer time in an ad consumed by slow motion improves the overall effectiveness of the ad.
- **Context score:** Systematic scoring depends on how many of the elements contextually connected with the topic in the non-conscious are present in the ad.
- **Metaphor score:** Scoring depends on the semiotics and imagery of the emergent or dominant metaphor identified and connected with the topic is present in the ad.
- **Brand semiotics score:** Scoring depends on whether the semiotics of brand personality and brand attributes are present in the ad.

These scores transform the way copy testing is performed and can break the ad into elements that lend themselves to factor analysis to determine which factors might be better suited for a brand or a category.

> *By 2018, an estimated 2 million employees will be required to wear health and fitness tracking devices on the job.*
>
> – Heather Pemberton Levy, "Gartner Predicts Our Digital Future," Gartner, October 6, 2015

Capitalizing on Fading Fads and Micro Trends That Appear and Then Disappear

A micro trend is a social or cultural or economic trend that appears and is guaranteed to disappear. These trends usually manifest themselves as having low resonance in the non-conscious as evidenced by low support in the media consumed by the non-conscious, and high levels of resonance in the conscious output of humans. Think about micro trends as things people don't feel very deeply about, but like to talk a lot about. Be they product ideas, or creative ideas, they tend to appear, and then disappear as quickly as they appear.

Sometimes the marketing impact is limited to a specific geographical region, and other times may relate to one or more particular demographic segments.

There is value to be had in exploiting these trends. Exploiting these trends serves a very unique and single purpose – creating the perception that the brand is "current" and "trendy."

These trends are not exploited through large commitments of capital through television advertising, but through small bets made primarily through social media in creating “tongue in cheek” tweets, and memes from the social media response teams.

Algorithmic extraction, and identification of trends with template responses, can become a vital element of the social tool kit of a company.

Even a well-regarded micro trend can disappear, such as the Unicorn Frappuccino from Starbucks, which enjoyed viral popularity on social media and gave rise to other original Starbucks drinks such as the Zombie Frappuccino. Or the game called Pokemon Go, which enjoyed worldwide popularity back in 2016, but is now out of fashion.

Capitalizing on Past Trends and Blasts from the Past

A trend or a metaphor that is past is one that has disappeared from the collective non-conscious and the conscious. These usually manifest themselves as having low resonance in the non-conscious as evidenced by low support in the media consumed by the non-conscious, and low levels of resonance in the conscious output of humans. Think about these trends as things people don’t feel very deeply about, and don’t talk a lot about. Be they product ideas, or creative ideas, they are usually things to avoid.

Sometimes the marketing impact is limited to a specific geographical region, and other times may relate to one or more particular demographic segments.

There is value to be had in understanding these trends that have truly passed by. First off, they serve as a vital diagnostic tool to ensure that no product or creative resources are stacked against things that have passed. Managing marketing and creative budgets first requires a walk away from assets that are truly not current and belong to the past.

Some trends that have passed do make a comeback later in a slightly different form. It therefore is vital to keep an eye on them to see when they are starting to make a comeback and become emergent again. This allows a company to get ahead of the product innovation curve.

Algorithmic tracking of past trends to identify when they resurface can be a critical component of the product innovation and creative innovation of a company.

RFP Response and B2B Blending News and Trends with Stories

Creating a request for proposal (RFP), or a sales pitch, is a common B2B requirement.

> **Step 1:** Algorithmically identify the metaphors embedded in the semiotics of an RFP. This gets at the heart of what the creators of the RFP are truly seeking. A way to algorithmically get at the problem is to identify as hypotheses a few hundred metaphors, and from the semiotics present in the RFP to rank and rate the top metaphors that are contained implicitly.
>
> **Step 2:** Make the metaphors explicit, and identify imagery and full semantic structure and frames that contain the metaphor.
>
> **Step 3:** Treating the RFP as a corpus, identify the contextual elements present in the RFP – tensions, daily activities, work, and occasions, to name a few.
>
> **Step 4:** Weave the contextual elements into the narrative of the response.

There are more than a few ways to find news stories and trends to report on in a storytelling style. They do it all the time in B2C circles. However, integrating news or trends into stories can be even more useful in B2B marketing. Recent B2B marketing trends that marketing professionals can use to their advantage include a keener focus on micro moments, a change in infographic design, email newsletters replacing brand magazines, and a change in copywriting style, among others. One current marketing trend is that B2B copywriting is now styled more than ever in a storytelling fashion. Writers must produce content that answers questions, as opposed to just dryly pontificating, or simply utilizing profitable keywords.

Sales and Relationship Management

Understanding a client is typically perceived as knowing their preferences for sports, religion, the number of kids they want, movies and songs they love, family status, the types of joke they like, and even their birthdays. This is really an old way of understanding the person.

The data structure of a relationship is defined as the metaphors that represent the non-conscious mind of a person along with the contextual elements pertaining to the relationships that are deeply resonant in their mind.

Knowing the topic of discussion, extract the dominant and emergent metaphors in the non-conscious mind of the client through algorithmic parsing of the media they have consumed. Activate the semiotics and imagery of that metaphor in conversations with the client.

Knowing the topic of discussion, identify the relevant contextual elements in the media consumed by the non-conscious mind of the client. Activate the contextual elements in conversations with the client.

Programmatic Creative Storytelling

Programmatic storytelling is in its infancy, and is anchored by algorithmic selection of relevant metaphors, relevant contexts, and five different story lines based on the Big 5 personality traits.

Data structure of a user comprises the choice of personality, the choice of metaphor, and the choice of relevant contexts. The algorithmic story generator generates the story in real time with the product or a call for action being highlighted. The template follows the typical template of a 30-second ad.

Template for Programmatic Storytelling

1. Opening metaphor use semiotics and imagery to connect to the non-conscious mind and the semiotics of the personality type
2. Emotional components of story and message using identified contextual elements that are personalized
3. Functional components of message using identified contextual elements that are personalized
4. Emotional call for action in story using identified contextual elements that are personalized, and the semiotics of the personality type
5. Closing metaphor with the brand

14
The Future of AI-enabled Marketing, and Planning for It

> *The popular definition of artificial intelligence research means designing computers that think as people do, and who needs that? There is no commercial reason to duplicate human thought because there is no market for electronic people, although it might be nice if everyone could have a maid and butler. There are plenty of organic people, and computer vendors can't compete with the modern low-cost technology used in making people.*
>
> – William A. Taylor, *What Every Engineer Should Know About Artificial Intelligence* (MIT Press, 1988)

Imagine a marketing function in which:

- Its population comprises computer scientists, cognitive linguists, neuroscientists, and algorithm developers
- Science and technology account for 70% of the budget
- Media buying and selling is done the same way high frequency trading happens
- Idea Exchange marketplaces flourish
- The garden variety is replaced by everyday masterpieces
- Consumer attention is for three seconds and becomes the most valuable commodity on the planet
- Expert systems capture the knowledge and failure of a billion campaigns the same way an infinity of chess games is stored in chess playing machines
- Rule-based decision trees and hierarchies of brand and product campaigns become the proprietary IP and the most jealously guarded trade secrets of marketing departments
- Algorithmic development of marketing causes companies to pull most of the services in-house to guard their secrets
- The generation of the core ideas for campaigns stay in-house, simple tasks of execution and production are farmed out

- Marketing expands – internal marketing to employees within the company through HR, and marketing to analysts and observers through the Investor Relations division of the company
- Agencies that deeply absorb AI and ML into their DNA become integral partners of large corporations

What Does This Mean for Strategy?

Successful businesses increasingly will need an AI and ML master strategy. Much as the introduction of computers, and subsequently PCs, into the business world altered that world in so many fundamental ways, and created the need for corporate IT strategies, so too will the adoption of AI and ML-based resources and methodologies into today's and tomorrow's businesses. Strategy is data driven, heuristics driven, and expert driven – there will be an increasing proliferation of companies that typically operate in the strategy and IT space into the marketing and market research space. The likes of McKinsey, Booz Allen, and BCG doing marketing strategy will increase as they find access to marketing data and paradigms through Machine Learning and Artificial Intelligence. Typical Operational Excellence, and IT implementation companies like Ernst & Young and Deloitte will find marketing and product innovation to be fertile grounds through ML and AI. This means that traditional market research companies will either have to innovate to stay ahead, or will find themselves in the sidelines of the games of evolution that occur as we speak in the marketplace.

What to Do In-house and What to Outsource

In terms of vast quantity and sheer complexity, data has become virtually impossible for mere humans to manage, without the help of sophisticated ML systems that process the information and learn from it. That is why an AI strategy is crucial to a company's success. Fortunately, the task of data analysis can be outsourced to a data science company.

When deciding which tasks to keep in-house and which tasks to outsource, key considerations include identifying which roles on your marketing team will be transformed by AI, and how the team member can adjust to the change. In reality, very few occupations are, or will ever become, *entirely* automated.

What India, TCS, and Infosys did to the world of IT and business process outsourcing is on its way to the world of marketing as well. Expect outsourcers to come knocking offering products and services at one-third of what they cost today. Moore's law is about to pay a visit to the world of marketing.

What Kind of Partnerships and the Shifting Landscapes

It is predicted that very soon, machines will manage 85% of customer interactions without any need for human input. The same may soon be true of employee interactions. It is expected that in the not too distant future, human workers will be partnering with chatbots. Chatbots will become a common feature in the workplace, and will be capable of offering coaching, training, brainstorming, and many other AI services to human employees.

Purchase data will continue to play a vital role as the primary source for secondary analysis. Data generators such as Amazon, Home Depot, and Walmart will have the potential to transform data into products, thereby competing with manufacturers; transform data into creative marketing campaigns, thereby competing with advertisers; and transform data into content, thereby competing with entertainers. The convergence of product innovation, creative advertising, and sheer entertainment is at hand. This singularity is a beautiful moment, of great economic promise for some, and utter disaster for others.

What Are Implications for Hiring and Talent Retention, and HR?

It is predicted that by 2022, at least 20% of workers will have some type of AI system that functions as a coworker. Job roles will change as automation takes its place in the marketing sector. This means that marketing professionals will need new sets of skills to stay competitive. Marketers who have education and/or training in science, technology, engineering, and math will become more valuable than ever before.

HR executives will soon find that they will need to understand the following topics to know how best to motivate a millennial workforce. They will realize that the metaphors – dominant and emerging – will drive the imagery, symbols, and language used in all internal communication. The successful HR organization will uncover these metaphors and expressions of these metaphors and will utilize them systematically in its day-to-day interactions.

- **Work** – its meaning and implications
- **Success** – what does it mean, and what are its metrics
- **Fulfillment** – what it is to be fulfilled, and how work facilitates it
- **Excitement** – the facets and aspects of its manifestation at work
- **Compassion** – what it is to experience and express through work
- **Empathy** – what it means in daily interactions and organizations
- **Human** – what being human means at work
- **Progress** – what it means and how it is measured, and fits into a worldview
- **Purpose** – how it gives meaning and drives motivation at work
- **Leadership** – what it means to lead, and to follow
- **Rules** – what they mean, and how compliance and the spirit of rebellion can coexist

Machine Learning will parametrize facets of these dimensions and will systematically learn as well the composite kind of person who succeeds in an organization. Factor analysis and simple PCA will help determine key hiring processes, which will be simplified through scenario generation and gaming to understand the factors that matter the most.

ML paradigms will cross industry sectors to penetrate areas never accessed before. For instance, in the world of Smart Beta, a tool in financial investing, there are factors that are used in understanding stock and portfolio performance. The most commonly used factors that have been drawn from academic research are the following.

- **Market:** *Stocks have greater return than risk-free securities.* There is an easy HR interpretation – employees provide better returns than risk-free consultants or contractors – so what is the right balance of employment and outsourcing?
- **Size:** *Small-company stocks have greater return than large-company stocks.* Would this apply to college degrees? That is to say, would a motivated candidate from a lower-ranked school perform better than an Ivy League candidate?
- **Momentum:** *Stocks that performed well recently have a greater return than those that did not do as well.* As high performance teams are put together,

how should recent performance of candidates factor into the mix?

- **Value:** *Cheaper stocks have greater return than expensive stocks.* Is it better to periodically restructure and restock the organization with lower cost employees?
- **Volatility:** *Stocks with lower volatility have greater risk-adjusted returns than higher-volatility stocks.* Is it better to score and monitor employees on emotional volatility, and continue to build organizations with emotionally stable individuals rather than mercurial geniuses?
- **Credit:** *Lower-credit quality bonds have greater return than higher-credit quality bonds.* Do employees with lower performance appraisals have a place in the overall portfolio of an organization so long as they are focused on specific tasks?
- **Term:** *Longer-term maturing bonds have greater return than shorter-term maturing bonds.* Should compensation mechanisms be biased toward longer-term gains rather than short-term performance incentives?
- **Quality:** *Stocks of companies with stronger profitability and stable income have greater returns than their counterparts*. What are the “basics” employees should have to enable them to be superior at work?

Note that while there are many smart and talented HR consulting companies, and they might have unique answers to all of the above questions, ML and AI paradigms naturally surface these analogies from one domain to another. It is easy to set up data science programs within the HR function that use factor analysis to unravel these vexing mysteries for each organization. Answers for one organization may be different than the answers for another organization, and this opens up an entire field of ML- and AI-driven HR consulting.

What Does Human Supervision Mean in the Age of the Algorithm and Machine Learning?

A good way to think about the need for human supervision of AI algorithms is by way of a famous thought experiment, which basically goes like this: if you simply program an all-powerful AI bot with the less-than-detailed instructions to “make paperclips,” the unconstrained AI function will do just that, only that, and nothing else but that, eventually transforming all resources on Earth (including us!) into paperclips.

The moral of the story is that detailed and ongoing communication between humans and robots is key to the success of AI, not to mention the safety of the human race.

Although human supervision is a definite requirement for AI in the foreseeable future, AI scientists still envision a world entirely free of it. After all, not only can an AI platform be programmed to learn, but it can also be programmed to adapt, and to use human-like reasoning in the decision-making process. It is even conceivable for an AI program to learn abstract concepts such as morality, ethics, and justice.

However, human supervision is still needed for many reasons, a main one being troubleshooting. For example, it is reported that Facebook recently had to close one of its AI labs when the chatbots began communicating in a language that was unintelligible to the researchers. Whether this is just romanticized dramatic reporting or simply gibberish that results from simple coding errors will remain unknown. Two of the biggest risks associated with AI systems include hacking and corrupted or incomplete data. Human intervention is needed to solve these problems.

Most importantly, an AI system needs to be strong enough to override sinister or dangerous commands. AI must be able to predictively analyze all the possibilities of a situation to determine if a command seems suspicious. If the command is determined to be suspicious, the AI program must have the flexibility to simply refuse the order, and block it from being executed.

How to Question the Algorithm and Know When to Pull the Plug

In 2016, Microsoft introduced a humanoid AI robot named Tay. Tay had all the looks and charm of an innocent, eager-to-learn teenage girl, tweeting pleasing comments such as “Humans are super cool!”

Unfortunately, less than 24 hours after going online, a gang of hackers found a way to exploit Tay’s commenting skills, forcing her to respond in inappropriate ways.

Tay only lived a day.

She was dismantled after posting the following series of tweets:

> “Bush did 9/11 and Hitler would have done a better job than the monkey we

> have got now; Hitler was right I hate the Jews; Repeat after me: Hitler did nothing wrong; Ted Cruz is the Cuban Hitler. That's what I've heard so many others say; Donald Trump is the only hope we've got."

And so it was that Tay became the first AI robot to be fired from a job for posting offensive and inflammatory comments on social media. This is a rather extreme example of knowing when to pull the plug on an AI algorithm, or at least knowing when to make some major adjustments to it.

AI algorithms must be constantly questioned, evaluated, and updated by humans, before, during, and forever after they are created.

In the same manner, marketers will need to monitor and alter AI algorithms in order to keep pace with changes in the marketplace, find flaws, make adjustments, and implement improvements in performance levels, *while* the campaign is in progress.

In fact, it is even possible to create a diagnostic algorithm that poses the right questions to test the health of another algorithm. But even this would require some form of human supervision.

Next Generation of Marketers – Who Are They, and How to Spot Them

Some members of this next generation of high tech marketing professionals will not be hard to spot. They will be the AI majors and data science majors of top universities, local colleges, and even engineering boot camps. They can also be found at any of a number of newly established software engineering schools for women.

But there will be plenty of room for smart, imaginative, adaptable people who may not possess elaborate technical skills per se, but who bring creativity and an innate drive to innovate to their jobs. AI and ML tools can deliver an immense array of data, in every conceivable format. Finding the "needle in the haystack" will remain something that some marketing professionals will excel at.

How Budgets and Planning Will Change

The purpose of using AI for budgeting, planning, and forecasting to achieve a company's overall and financial goals is better accuracy, greater speed, and lower overall costs. Examining how these activities are influenced by AI can illustrate how much of a company’s budget should be invested in the development of a smart corporate AI strategy.

15
Next-Generation Creative and Research Agency Models

> *The coming of computers with true humanlike reasoning remains decades in the future, but when the moment of "artificial general intelligence" arrives, the pause will be brief. Once artificial minds achieve the equivalence of the average human IQ of 100, the next step will be machines with an IQ of 500, and then 5,000. We don't have the vaguest idea what an IQ of 5,000 would mean. And in time, we will build such machines – which will be unlikely to see much difference between humans and houseplants.*
>
> – David Gelernter (attributed), "Artificial Intelligence Isn't the Scary Future. It's the Amazing Present," *Chicago Tribune*, January 1, 2017

Imagine an agency where:

- Algorithms extract metaphors that are relevant to any given topic or category – what is "Natural" a metaphor for, and what is "Ice Cream" a metaphor for.
- Algorithms extract and classify the metaphors into things that are emergent metaphors where the social zeitgeist is just beginning to coalesce.
- Algorithms identify and extract semiotics, words, concepts, and imagery that automatically bring the metaphor to life.
- Algorithms extract sentences from a wide body of literature that contain the category, brand, product, and the identified semiotics – thus forming the basis of word-based descriptions and product definitions.
- Algorithms automatically measure the distance of the descriptions from the brand personality, and-fine tune them to be closer.
- Algorithms automatically extract facets of the creative in non-conscious minds – be they cultural tensions, life pressures, social issues, rituals, and so on, connected to the topic, category, brand, and product.
- Algorithms extract story structures that are relevant to the topic.
- Algorithms transform standard 2-D images into 3-D interactive augmented reality scenes.

- Algorithms create context-sensitive digital overlays on demand.

Creative execution begins after this, and leverages what algorithms have identified as ingredients in the non-conscious mind that are already connected to the topic and category.

By definition, the work of traditional market research and advertising agencies has always been rooted in nineteenth- and twentieth-century labor models. Which is to say, the work performed by people. From the initial strategies and tactics developed to the finished end product, these professions, like many others, have depended on – and thrived from! – the fruits of the human mind.

It is no exaggeration to say that AI and ML are hard at work changing that basic premise today – just as they are in so many other fields. But it is a mistake to take that remarkable proposition and extend it to conclude that people will become peripheral to the tasks at hand. Certainly, automation of many labor-intensive types of work is inevitable. AI- and ML-powered systems can simply do the work faster, and arguably better (if you accept the idea that "better" means more accurately and with far-greater resources applied), so naturally they will preempt human labor because they make more economic sense to use.

But at the close of this book it behooves us to remind ourselves that while AI and ML are truly game-changing advances, they should serve us, and not the other way around. Perhaps it is oversimplifying to say that they are tools – highly powerful, sophisticated tools, to be sure – but in the fields of market research and marketing agencies, that is essentially how they should be viewed.

If the goal and the benefits of marketing research revolve around the gathering and analysis of useful data that can be applied to better understand consumers and their behavior . . . and if the goal and the benefits of marketing services companies (broadly stated) revolve around creating methods and messages designed to inform, persuade, and motivate consumers . . . then the human element should always be at the center of the work.

The experts who are building the next generation of advanced AI and ML platforms designed specifically for helping drive innovations in new products and marketing are doing so with the view that the "work product" their systems deliver are intended to enable professionals to do their jobs better, smarter, more efficiently, and more effectively. These systems are designed to provide superior bodies of information. Superior guides to greater inspiration. Superior avenues toward achieving new products and features that consumers will innately find attractive and responsive to their needs – felt, and unfelt just yet. Superior resources to spark the creative “aha!” moments that lie at the heart of the most successful marketing messages.

> *Business leaders said they believe AI is going to be fundamental in the future. In fact, 72% termed it a “business advantage.”*
>
> –“How Artificial Intelligence Is Pushing Man and Machine Closer Together,” PwC, April 2017

What Does an ML- and AI-enabled Market Research or Marketing Services Agency Look Like?

An AI-enabled research agency looks remarkably similar to those of today: an office full of computers, where people are feeding data in, then retrieving and analyzing useful patterns of information. A key difference is that, unseen, powerful AI and ML systems are unleashed to scour multiple databases at speeds and with such analytical prowess that no human or team of humans could match them. A very key difference in the type of work done in this “new” market research company is that staffers are now giving directions to these systems, and refining their output continuously.

One of the most useful aspects of AI and ML for market research and creative purposes is that they can point us in new and perhaps different directions. On the best days, at the heart of both industries is that “aha!” moment – the point when, unexpectedly, a data point, a partial song lyric, an image, a metaphor, or any number of other “stimuli” are delivered by the system. That can, and hopefully will be, the departure point for the experts who work there. Those facts, those insights, those inspirations are what can lead to true (better) understandings and innovations in the marketplace.

Another difference in the new models of market research and marketing services agencies will be the presence of a new tier of staffers. They will be people who are schooled in the basic architecture and operations of AI and ML systems. In the very largest firms, there will likely be data scientists who can help gather and distill the output that AI and ML systems will deliver.

Creatives will increasingly look “outward” for ideas and inspirations; meaning they will depend a bit less on individual or team brainstorming sessions and more on directing their AI and ML systems to go out and gather everything they can find that might be relevant and applicable to the marketing challenge clients are facing.

The defining aspect of the agency is – rapid iteration. The ability of this agency to continuously iterate creative that singly breaks the dysfunctional world of checkpoints and copy testing.

Advertising will increasingly shift towards experience creation – passive entertainment will migrate to active, and interactive entertainment and brand storytelling. Augmented reality, and augmented decision-making will enable consumers to enter and be present in virtual environments of their creation.

What an ML- and AI-enabled Research Agency or Marketing Services Company Can Do That Traditional Agencies Cannot Do

An AI- and ML-based agency will be able to leverage the bits and pieces of creative already embedded in the non-conscious human mind. As outlined above, AI- and ML-enabled research and marketing services agencies will differ from their traditional predecessors in a number of important ways. Being able to generate insights in real time, and being able to creatively act on them in real time is another big difference. High frequency trading enables computers to convert miniscule differences in equity prices into profit. Algorithmic creative will similarly convert miniscule differences in cultural nuances into brand awareness and purchase intent. The integration of AI and ML systems will enable research and marketing professionals to make better informed, more insightful, and more impactful decisions at a much faster pace.

An AI- and ML-enabled market research or marketing services agency also lets marketers look into the future and run models to help clients plan their business strategies, new product development processes, and creative development paths. Again, the speed and flexibility with which these processes can operate far exceeds anything that is possible with legacy systems and procedures.

RAD JAD – Rapid Advertising Development and Joint Advertising Development – methodologies will start to proliferate. The RAD JAD methodology – first called Rapid Application Development and Joint Application Development – are core methods for rapid prototyping and user testing of software. Paradigms from the world of software development and user interface development will shift into the world of advertising and interactive consumer experience creation.

The non-conscious mind functions like the Github for Marketing. It already has bits and pieces of code that have been worked out and proven. All the agency has to do is to assemble them in different ways to create fundamentally different designs.

> *Forty-seven percent of digitally mature organizations, or those that have advanced digital practices, said they have a defined AI strategy.*
>
> – "Fifteen Mind-blowing Stats about Artificial Intelligence," https://www.adobe.com/insights/15-stats-about-artificial-intelligence.html

The New Nature of Partnership

As AI and ML resources – both in-house and external – increasingly penetrate almost every business of any size, globally, the interactions between "client" and "vendor" will evolve.

Ad agencies especially have been enamored of thinking of and referring to themselves as "marketing partners" with their clients. That concept has lost ground in recent decades, for a variety of reasons. But with the advent of AI and ML systems, that idea of "partnering" toward marketplace success will likely be reborn.

The swift sharing of information and insights . . . the speedy implementation of ideas and learning into new products and new messaging campaigns . . . and the ability to adapt to changing marketplace conditions with far greater agility will be the universal hallmarks of these client/agency partnerships.

Clients will benefit from knowing that their marketing campaigns are derived from far larger and superior reservoirs of information and insights. Agencies will benefit from clients having more overall confidence in the creative solutions offered, and from somewhat faster and more assured client decision-making about budgets and choices in the production process.

The key disruption is the arrival of strategic consultants like McKinsey and the likes of Accenture and Deloitte to the world of creative. Unleashing software-enabled design, data-driven creativity, and app-driven consumer experience creation will change the nature of this partnership.

> *For those enterprises already in the AI fray, top-performing companies said they are more than twice as likely as their peers to be using the technology for marketing (28% vs. 12%). Unsurprisingly, analysis of data is a key AI focus for businesses, with on-site personalization the second most commonly cited use case for AI.*

– "Digital Trends," Adobe, 2018

Is There a Role for a CES or Cannes-like Event for AI and ML Algorithms and Artificial Intelligence Programs?

Awards serve several purposes, and as a species we seem to love them. So the question begs: With AI and ML advancing at such a rapid rate, and new algorithms and digital innovations being created constantly, does it makes sense to stage an event where the latest AI and ML ideas and products can be showcased and celebrated?

Why not? We live in an age when technology drives most aspects of our daily lives, and where tech icons are idolized. Such an event could keep us better informed about the amazing progress made in these fast-moving fields, and provide ideas and guidance about how best to apply these hyper-intelligent tools to achieve the best results.

It might also serve to help allay some of our concerns about the potential rise of "super machines" that theoretically could threaten our freedom, or even our existence. Recognizing outstanding achievements in AI and ML that help improve our lives, through granting better access to the most useful information we seek, better performing and more relevant products and services, more engaging entertainment, and other positive outcomes seems worthwhile.

Now all we need is an AI/ML program to design the ceremony, create the award, process all the entries from around the world – and perhaps even select the winner!

Challenges and Solutions

Digital products and services connect us not only to each other and information and entertainment sources, but also to a hitherto-inconceivable range of consumer products and services. AI and ML power the internet's abilities to understand (track) our current needs and interests, deliver things that respond to those desires, and also infer from them what we might want or enjoy next.

The issues surrounding this data collection and usage are the focus of much debate, and disagreement. Post-argument, though, are the foundational platforms that we rely on to an ever-growing degree worldwide. Whether it is social media outlets, online retailers, or entertainment sources, the fact is that AI and ML are already at work driving the engines that get us what we want, when we want it. And they suggest to us other products and services, or information and entertainment providers, we may find worthwhile.

The challenges this poses to businesses large and small are becoming geometric in growth and complexity. Product life cycles are narrowing. Product quality is becoming more and more standardized, at higher and higher levels (think automobiles, computers, software, medical technology, entertainment, and so many more categories). Product differentiation is becoming more difficult to achieve and maintain as a long-term competitive advantage. Pricing is under pressure. Consumer expectations and demands are becoming more sophisticated and focused.

At the same time that global competition is heating up at inexorable rates, the global marketplace is splintering in terms of societal pressures and governmental regulation. What's permissible and acceptable on a social media platform here, may well run up against opposing cultural and legal obstacles there.

On top of that, media are simultaneously fracturing and coalescing. News outlets have multiplied beyond measure, while "legacy" structures like traditional stand-alone Hollywood movie studios are narrowing in number, merging into conglomerates, and encountering an array of international competitors (China has recently built the world's largest and most technologically advanced studio), as well as digital purveyors from Amazon to Netflix.

All of these trends are impacting the marketing of consumer goods and services in ways that were largely unimagined just a few short years ago.

What can AI and ML contribute to addressing these central issues – and to offering solutions that meet both consumer and business needs?

The most basic answer is these twin technologies' abilities not just to see around the next corner – but instead to "fly over the mountain" and capture both short- and long-range insights and sources of inspiration that defy and far exceed humankind's core intellectual and instinctive resources.

To put it bluntly: we are simply not biomechanically designed to investigate, analyze, filter, incorporate, and ultimately synthesize the sheer scale of information that AI and ML systems are designed to do.

On the scary sci-fi front, this fact conjures up doomsday scenarios of "the machines taking over."

As titillating as the “end of humanity as we know it” storyline can be, it is not a foregone conclusion. What is a foregone conclusion is that AI and ML, in all of their near-infinite scope and variety of applications, are not only here, they are here to stay and are multiplying – again, geometrically – to form the unseen basis of much of human activity, products, and services as the century extends.

Acknowledging that there will inevitably be negative – perhaps evil – embodiments of AI and ML, the same can already be said for other advances in human history. Nuclear power can heat and light our homes, and it can also explode in lethal fireballs.

But this final chapter’s focus is on a couple of the “better angels” of our nature, when it comes to AI and ML being a positive and productive force. There are more profound examples that can be cited, outlining AI and ML’s contributions to science, medicine, knowledge access, and too many more areas to mention. But here are some thoughts to consider about these technologies’ contributions to, and impact on, creative elements in marketing, and market research.

The old adage that if you “give 100 monkeys typewriters and enough time, sooner or later a Shakespearean play will appear” – even if theoretically and arguably possible – has always been a far-fetched notion. Explaining, much less scientifically replicating, human creativity has always been more like capturing lightning in a bottle: if attempted, it’s likely to end badly.

It is conceivable (and predicted by some) that, at a point, AI and ML could become so advanced that for all intents and purposes, they are capable of imitating human creativity to a degree where the two are virtually indistinguishable from each other. Notice we said *imitating*. Driven by enough processing power and sophisticated supervised/unsupervised learning protocols, creative ideas and elements could be brought forward with little to no human interaction. These deliverables might prove to be “good” enough to serve certain needs without enhancement by human minds and hands. Especially when it comes to technology, “never say never” is a wise premise to follow.

But this is a somewhat dystopian view of the future. Unless and until that creative “singularity” is ever reached, the sheer magic of creative inspiration and its application will remain with and reside in the human mind.

That said: things *will* change in this realm of creative imagination and expression. In fact, they already are.

Software systems that can learn video editing have already produced segments where visual sequences, and specific styles and paces of cuts and visual effects, are respectable competitors to what a human editor might assemble. Could a master editor conjure a “better” version though? The answer can be argued endlessly. The point is: due to AI and ML, technology today can already perform a task that until now has relied exclusively on human skills and creative vision.

Storyboards, mood boards, music tracks, and more can all be automated to one extent or another. With access to vast and varied digital databases of images and sounds, and driven by expertly coded algorithms, AI and ML platforms can and will gain ever more ground in coming closer to what could be acceptable levels of creative quality compared to exclusively human-curated examples.

So are copywriters and art directors, editors, music composers, directors of photography, sound engineers, musicians, painters, set designers, costume designers, location scouts, and others approaching extinction as creative species?

No need to roll out the rocking chairs just yet.

AI and ML’s immediate promise and potential – and this will be true for the indeterminate future – is their ability to accomplish what would otherwise take a human to do, but so much faster . . . with such greater breadth and depth . . . accessing so much larger bodies of resources . . . and capable of assembling the findings more quickly and in ways that can be helpful and yes, even inspiring to the human observer of the process.

So perhaps the wisest way to view this emerging brave new world is to anticipate that first, human creative inspiration is not going to wither and die; and second, that the “creative function” will be *elevated*, not eliminated.

As useful, surprising, and arguably even brilliant as the end product of AI- and ML-based systems will likely become, the spark of human creativity will not only *not* be extinguished – it will be the final arbiter, the “god in the machine” if you will. The best creative minds will sow these systems with ideas and instructions, and reap what they want from them.

Just as online search engines may produce somewhat different results when posed with a slightly different query, so too will AI and ML systems when directed toward the production of creative ideas and materials.

What are the implications of this “elevated role” for creative professionals and creatively based industries? A useful comparison is the autonomous automobile. Yes, it can “drive itself.” But it still needs to be told where to go before it can take you there.

The different skill sets that will evolve over time will entail creative professionals understanding how AI and ML systems work, and how to employ them for the best results. Does this mean mastering coding as a prerequisite to becoming a successful twenty-first-century creative? Not necessarily. But understanding how to guide these systems, through “directed questioning” and database selection, will be a very valuable asset in the creative toolbox.

Big Data

The term “Big Data” is both overused and misunderstood today.

As remarkable as massive warehouses of digital data are, obviously in and of themselves they don’t do much. It’s terrific to have assembled petabytes of loyalty-card usage patterns and customer profiles – an accomplishment unto itself. But obviously, that data can’t “tell” you much all by itself, in terms of how best to put it to use to serve customers better, improve products and services, solve current and possible future problems, and gain a competitive advantage in the marketplace.

The same is true for the creative elements used in developing marketing materials. How wonderful is it to have instant access to millions of recorded songs? Very wonderful, until you have to find that one lyric, or those six unconnected songs, or that obscure international musical jingle that – with some human thought and creative inspiration – might be the key to a successful new marketing campaign.

In the most general terms, AI and ML are technologies capable of not just assembling those disparate elements for you. They can also take “one giant leap” forward and create something from them that might take you to the solution for that marketing challenge that your company, or client, are seeking.

Let’s begin at the beginning: *strategy.*

AI- and ML-powered Strategic Development

Few will argue that a successful product, service, or marketing effort can be achieved without a sound strategy at its core. Examples otherwise are the exception, not the rule.

How can AI and ML help arrive at smart, effective marketing strategies? By elevating the strategic development process to a higher level. The ability of AI and ML-driven systems to gather, compare, contrast, infer, and convey learning are orders of magnitudes more than anything the human brain can accommodate, much less accomplish.

To draw a crude parallel, it is along the lines of attempting to understand the cosmos through Galileo's Earth-based telescope, versus the view of the universe through the Hubble Space Telescope. Both can see stars. But only one can capture and analyze light from countless numbers of invisible sources billions of light years away.

Asking the right questions – directing AI and ML systems to go out and retrieve the most relevant and useful data – will be at the heart of the most sophisticated (read: effective) strategic development processes. Instead of sitting around a conference table crunching numbers, studying a competitor's strategies, and brainstorming possible solutions, the modern strategist will beckon his or her trusty AI/ML system to deliver findings that offer a much broader, varied, and unobtainable-by-any-other-means series of possible strategic paths.

In a very real and practical sense, strategists will now have an actual tool kit to perform their work. And it will be a tool kit with the most advanced, sophisticated tools in it.

Such learning may confirm the strategist's initial understanding, thoughts, and inclinations; in which case, a great deal of time, effort, and informed guesswork can be saved.

Or they may offer up an unexpected, and therefore unseen, set of other possibilities. "Surprise" may well be the best strategist's near-constant work companion – and that emotional/mental response will be warmly welcomed, because it's the appropriate response to a fresh and inspiring set of findings.

So the new skillset for business and marketing strategists – and for those in the fields of politics and other "soft" categories as well – will be a fundamental grasp of what AI and ML are in terms of their broad capabilities, and an increasingly knowledgeable sense of how to apply them in pursuit of the smartest strategy. Just as with a Google search today, the more these systems are used by someone, the "better" someone can become at employing them to achieve the most effective results.

Vendors in the field of market research, strategic development, and business consultancy will compete – and prosper – on the basis of the quality, depth, and breadth of their grasp of AI and ML technologies, and the demonstrated commercial success of their outputs for clients.

There may well be one more step in the search for the best strategy – and that may be the singular, brilliant “leap of logic” from the human mind that takes a strategy from the merely mundane to the magnificent. That does of course occur today, even without the digital dynamism that AI and ML deliver. But those authentic and extraordinary leaps are few and fitful. With AI/ML harnessed to the task, that “aha!” moment will become much more frequent, more easily obtained, and more actionable.

Creative Execution

“Actionable” is where AI and ML play their next role in the creative process as applied to marketing.

The twin technologies’ ability to search for, summon, and synthesize data from virtually unlimited digital sources sets them apart from any other method of creative development and execution. At one level, AI/ML deliver practically limitless variations on a given theme; set the systems going, and in very short order they will provide a proverbial fountain of inputs: songs, movies, TV shows, poems, speeches, artwork, trends in online inquiries on any specific subjects, favorite phrases in Swahili . . . you name it (ask for it, even in general terms), and the mind of the machine will circle the globe and retrieve it all for you.

For creative people seeking fresh inspiration, this can be the Fort Knox of ideas and expressions. Tweak the systems a bit, refine the parameters, and they will immediately procure a whole new set of stimuli. With machine learning at work, the system will refine itself, seamlessly fine-tuning to gain even more focused and relevant results.

At this level, the products of the search and analysis processes are simply there for the taking. Rummage through them, unearthing a visual nugget there, a partial song lyric there, an up-to-the-minute glimpse at what consumers are inquiring about online, and allow the human creative instinct to extract and apply what it will (obviously, not violating IP rights in the process, of course!).

Beam Me Up

But here is an excellent example of where the concept of an “elevated” creative process made possible with AI/ML can take hold.

Let's say the assemblage of bits and pieces of the stimuli listed earlier has been gone through, anywhere from superficially to an in-depth dive into that nearly bottomless ocean. Let's further say that the restless drive for creative inspiration has not yet been fully satisfied – it just feels like there may be something more, something powerful, something that would take a commercial or a campaign to a whole other, higher level.

With the right algorithms, you hold the keys to a creative universe essentially of your own making. Intelligent systems unleashed in this way will race headlong to fracture and recompose, combine elements together in ways that would take a million human minds a million years to complete, and find entirely unexpected connections that elude the conscious human brain.

Ultimately, combined with CGI/VR/AR capabilities, a "finished" end product could be delivered. Don't like the aircraft image in the fifth scene of the spot? Task the system to make it different.

Want the music bed of that radio commercial to be less bass-y, more upbeat? Cue the algorithms and then listen to what you heard in your head.

As we approach the age when holograms become mainstream parts of our visual environment, the opportunities that AI/ML will present for using that technology for marketing purposes will become manifest. Near-lifelike individuals appearing at your beckoning, addressing you by name, entertaining you, reminding you, educating you, encouraging you . . . the possibilities tempt the imagination.

Neuroscience teaches that the non-conscious mind responds quickly, strongly, and positively to "personalization." Starbucks figured that out intuitively – their ordering process involves writing your own name on the cardboard cup. Not only does that ensure that you get the nonfat mocha cappuccino that you want, but it also has another, invisible effect: it satisfies the subconscious. It makes you feel recognized – and that in turn "binds" you more effectively with the whole Starbucks experience.

Project that same effect into your living room, when George Clooney appears before you at your beckoning, to tell you – by your first name – all about the latest Mercedes model that you're interested in. AI and ML are at work here, not only tracking your expressed interest in a new Benz, but also triggering George's image and programming the hologram to personalize the message.

Follow that with a VR test drive, all in the comfort of your own home, and a visual rundown of all the choices available to you in terms of color and options.

Of course, you'll be able to order at your will as well.

Will Retail Be a Remnant?

This example begs the larger question of these technologies' impact on brick-and-mortar establishments, and the economic and societal impacts that will flow from that. Amazon has already revolutionized the shopping experience. Apple and Netflix have done the same with entertainment. Other examples abound, in a host of categories.

AI and ML will effectively "bring the world to you" in much the same sense as digital technology overall brings us the world at our fingertips. The key difference is the power that AI and ML systems have to customize your experience, refine offerings to whatever degree you wish, render aural and visual (and ultimately tactile) stimuli that actively engage you, respond to your needs and desires increasingly intuitively, and generally render unto you an alternative, parallel environment to the real world. We've all seen enough sci-fi movies to understand that, over time as the technology becomes ever more sophisticated (and yes, powerful), these two worlds will become increasingly indistinct from each other.

Human nature rules, however. It will no doubt continue to even in this brave new world. We are tribal. We crave human interaction, much as we seek information and entertainment and physical and emotional sustenance. So AI and ML systems will bend to those drives.

Car dealerships are likely to morph from their traditional showroom orientation, to a hybrid of local physical service centers and VR sales methods. Futurists have already predicted the shift from the old-school department store to new models where large-scale entertainment is the centerpiece, and personalization at the POS is keyed to individual needs and interests. Movie theaters will be much more fully immersive multisensory experiences, tailored even more to the "blockbuster" phenomenon we enjoy today.

Getting Real

Marketing in the age of omnipresent AI and ML systems will not only benefit from the powerful potential these systems offer and deliver – it will also face very real challenges.

The human brain is exquisitely attuned to what is “real” and what is not. We evolved that way because such discernment could mean the difference between survival and extinction. Even as immersive as some gaming and entertainment products are today, the non-conscious mind still draws the clear distinction between what it “knows” to be genuine, versus what it perceives as “manufactured.”

AI- and ML-powered processes and products will have to address that eternal truth about the human mind, and strive mightily to achieve the delivery of stimuli that the non-conscious mind will accept as close enough to real to be virtually indistinguishable.

Over time, our brains may adapt to VR worlds to such an extent that they become literally indistinct from reality. A small indication of that kind of adaptability is the pace of visual stimuli today, versus the pace found in the early 20th century. Our visual systems were already capable of processing data at very high/fast rates – all it took was for delivery systems (film, television, and digital) to catch up and deliver data at those higher speeds. Today, we take that for granted – the proof being that watching “old” forms of visual information and entertainment now seems hopelessly slow and labored.

It Begins – and Ends – with an “A” Word

What will it take for marketing to succeed in the Age of AI and ML?

Technological infrastructure is one starting place. The basic assumption is that technological advances in terms of hardware and software capabilities, digital transmission speeds, and related factors will continue and accelerate. Rendering stimuli and environments that will truly engage and motivate discerning consumers is a tall order, but all indications are that we will approach and achieve that systems standard sooner rather than later.

But the rendering of marketing messages and materials that will also win consumers’ attention, emotional engagement, and memory retention poses its own formidable challenges. Coming back to the Big Data point made earlier, having the capacity to do something is not the same as creating methodologies, taxonomies, and skillsets to accomplish something truly worthwhile and effective with it.

Marketing professionals will need to master more in this new age. A fundamental grasp of what algorithmically driven systems are, what they can and cannot do, and most importantly how best to marshal all of the resources necessary to explore and exploit their capabilities, is going to be the new norm.

Business schools will need fundamentally to revisit and revise their current curriculums. Core competency in AI and ML systems and best practices will be “baked into” B-schools’ course offerings, and across more than just marketing. Case studies will bore into how finely tuned algorithms helped drive successful business, product and service, and marketing strategies. (This extends to finance majors as well – AI- and ML-driven investment decisions and products will be even more of a central component in boardrooms and on Wall Street than they are today).

Marketing service companies will undergo similar seismic shifts. Consultancies will compete for (and with) in-corporation talent with well-honed AI and ML skills and know-how. What we think of today as advertising agencies will have to become much more holistic in their strategic and executional approaches to creating and crafting marketing materials. Again, a basic but also sophisticated grasp of AI and ML systems will be the starting point.

Media buying will move from today’s increasingly programmatic-buying technologies to a multidimensional, matrix approach that will factor in more than demographics, viewing, and purchasing patterns. Relying on AI and ML-based systems, media buyers will be “looking over the mountain” as well as at the immediate road ahead, in order to anticipate marketplace and media trends that will arrive at far faster paces than before.

The smartest, most tech-savvy suppliers of information and entertainment will welcome the near real-time feedback from viewers, powered by AI/ML systems, that enable them to adapt their products to respond quickly and accurately to that feedback.

Compensation across many sectors of the business and marketing realm will become more variable and less fixed than today. Fees will ride (on a real-time basis) on the measurable marketplace success achieved by AI and ML systems. Those suppliers with “better” algorithms, more expansive and sophisticated methodologies, smarter and more adept staff, and swifter adaptation skills will prevail and earn the compensatory rewards that reflect those advantages.

An apt, if somewhat incomplete, analogy is a season-winning Formula 1 racecar team. The company with the best/most effective technologies, the most well-trained and motivated pit crew, and a smart and highly skilled set of drivers handling it all have the highest likelihood of succeeding.

It is – and will increasingly be – the core algorithms at the heart of an enterprise that drive it to the finish line ahead of the pack.

But it is *not* only those core algorithms that will be the decisive factors in achieving that win.

It will be the particular vision of how to combine algorithmically driven technologies, with digital resources, with a deep and meaningful understanding of core neurobehavorial science, with a set of gifted individuals who master them all and make the resulting combination produce something extraordinary: something that not only looks over the mountain – but moves it as well.

It begins with algorithms.

It ends with that glorious, mysterious, elusive, magical moment of “aha!”

ABOUT THE AUTHORS

A.K. Pradeep

Dr. Pradeep is a leading international business consultant and entrepreneur, having launched a series of successful firms across a number of categories. Most recently, he is the founder and CEO of MachineVantage, an advanced high technology enterprise specializing in the application of Artificial Intelligence and Machine Learning for innovation in product development and marketing.

Earlier he founded and is the chairman of Smilables, a computer wearables company designed to revolutionize baby brain development over the critical first two years of life.

Dr. Pradeep also created and served as the CEO of NeuroFocus, which ranks as the world leader in the consumer neuroscience field, with numerous patents for its breakthrough technologies. Following the acquisition of NeuroFocus by Nielsen, Dr. Pradeep served as chief provocateur of Nielsen.

Dr. Pradeep founded BoardVantage, which provides web-based corporate governance platforms for corporate boards of directors. The company was subsequently acquired by NASDAQ.

He was the founder and managing partner of Meridian Consulting LLC, a privately held firm specializing in governance consulting and customizing and applying General Electric’s best practices to multiple industry sectors.

At the outset of his career Dr. Pradeep was a scientist at GE Corporate Research and Development where he worked extensively with various global businesses, including medical technologies, satellite navigation systems, and other classified technologies.

Dr. Pradeep’s professional honors include the Great Minds Award from the Advertising Research Foundation, the organization’s top prize, which recognizes “an individual who brings excellence to advertising research in the category of research innovation.” He was also named Person of the Year by the USA India Business Summit for his “impressive innovations and achievements in the field of neuromarketing.” Dr. Pradeep has been featured in major media outlets including *ABC World News*, CNN, *The Economist*, the *New York Times*, and many more.

Dr. Pradeep holds over 50 US and international patents, has 100 pending, and has been published in a range of scientific journals. He is a frequent keynote speaker and panelist at domestic and international business conferences on issues of corporate strategy and marketing. His book *The Buying Brain: Secrets for Selling to the Subconscious Mind* (Wiley, 2010) is a best-seller published in multiple languages that offers an in-depth explanation of how the latest advances in neuroscience are impacting brands, products, packaging, in-store marketing, advertising, and entertainment content on a global basis. He is also the author of *Governance Beyond Sarbanes Oxley – Five Easy Pieces*, a guidebook for board members and senior corporate executives.

Dr. Pradeep earned his PhD in Engineering from the University of California at Berkeley.

Andrew Appel

Andrew Appel is president and CEO of IRI. Since joining the firm in June 2012, Appel has led IRI's transformation from a leading insights provider to delivering growth as a technology-focused Big Data company, which integrates the world's largest repository of otherwise disconnected data sets that lead to purchase on a cloud-based platform. Appel has overseen the development of a variety of products and solutions that leverage Big Data to deliver growth, including IRI's recently announced enhanced Liquid Data Platform, Shopper Marketing Cloud, and Prescriptive Analytics.

In addition, under Appel's leadership, IRI created a best-of-breed market research ecosystem with over 30 active partnerships, and significantly expanded its business into new industries and markets, including China and Australia. During his tenure, IRI has consistently grown revenue and EBITDA and now serves over 5,000 clients worldwide.

Prior to joining IRI, Appel held a number of senior leadership positions for Chicago-based professional service firms. As chief operating officer at Aon, Appel developed and implemented a successful revenue growth plan for the company. He also led the team responsible for the $5 billion acquisition and integration of Hewitt, solidifying Aon's leading position in human resource consulting and outsourcing. Before that, Appel was CEO of two of the firm's three global divisions: Aon Consulting, a leading benefits and human resources consulting firm, and Aon Benfield, the world's largest reinsurance broker. Under his leadership, the divisions achieved operational excellence and improved financial results. Following his time at Aon, Appel served as the senior vice president of Revenue Operations for Accretive Health.

Earlier in his career, Appel was a senior partner at McKinsey and Company, where for 15 years he advised leading global financial institutions on a wide array of operational, strategic, and organizational issues. Appel was also a founding member of McKinsey's Global Business Technology and Operations Practice, which he led for seven years and helped grow from 50 to 300 consultants.

Appel holds a bachelor's degree in economics from the University of California, Los Angeles, and an MBA from the University of Chicago, where he was the Henry Ford II Scholar. He is a board member of World Business Chicago, Alight Solutions, and has served on two task forces related to resolving the pension shortfalls of the state of Illinois and the city of Chicago. He and his family reside in the Chicagoland area.

Stan Sthanunathan

Stan Sthanunathan joined Unilever in July 2013 as executive vice president of Consumer & Market Insights. As chief provocateur, he heads up the insights function globally and is based in London.

Prior to joining Unilever, he was vice president of Marketing Strategy & Insights for The Coca-Cola Company in Atlanta, heading up the function on a global basis.

Some of his many external affiliations included being a board member at the Advertising Research Foundation (ARF). He was also the chairman of Online Data Quality Council of ARF, and has lectured on insights at MIT Sloan School of Management, Yale, and the University of Wisconsin. He is a patron and fellow of the Market Research Society.

INDEX

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

U

V

W

Y

Z