Aggregation Methods in the Wisdom of Crowds: A Literature Review

Thanh Trung Nguyen[1], Hai Bang Truong[2]

[1] Faculty of Information Technology, Ho Chi Minh City University of Foreign Languages and Information Technology, Vietnam
trungnt2@huflit.edu.vn
[2] Saigon International University, Ho Chi Minh City, Vietnam
truonghaibang@siu.edu.vn

Abstract. The phenomenon of the Wisdom of Crowd has been observed in several problem domains where the collective opinions of groups tend to be more accurate than those of individuals. In this context, the aggregation of individual opinions is a crucial factor that determines the success of unleashing the wisdom of the crowd. This paper aims to conduct a literature review on various aggregation methods employed in recent researches. We found that despite the increasing number of studies that deal with large and intricate datasets, conventional aggregation methods such as arithmetic, geometric, weighted aggregation, and mode continue to be widely used in the field of Wisdom of Crowds.

Keywords: Wisdom of Crowd, Aggregation methods, Arithmetic average, Geometric average, Weighted average.

1 Introduction

The concept of Wisdom of Crowds refers to the approach of the aggregation of multiple individual estimates to obtain a better collective one. Such an approach can outperform individuals, even domain experts, in various prediction and estimation tasks [1]. Surowiecki claims that a mathematical or statistical aggregation over the judgments of a group of individuals can be more accurate than those of the average individuals because of the benefit of error cancellation [1]. In his work, he describes four characteristics that make a crowd intelligent. First, the group should be diverse, as this allows for various individuals to complement each other by contributing unique pieces of information. Second, a decentralized structure is crucial for the group, without any centralized authority directing or influencing the answers of individuals. Third, it is also essential that the individuals within the crowd act independently of one another. Fourth, when the information of many individuals is pooled, they must be aggregated into a collective opinion, with numerical contributions and statistical methods often serving as the basis for aggregation. While diversity, independence, and decentralization are important factors, the aggregating method plays a key role in consolidating and synthesizing the individual opinions and knowledge within the crowd. It helps consolidate information, filter out noise, identify consensus, handle conflicting information and enhance collective accuracy.

The reason why it is possible to aggregate multiple predictions and yield a superior outcome compared to individual predictions can be attributed to the Law of Large Numbers [2]. According to the “many wrongs principle” [3], predictions are modeled as the truth plus a disturbance. The crowd wisdom is founded upon the aggregation of individually independent guesses, which feature random or symmetrically distributed errors. When numerous unbiased individuals make an estimate, the likelihood of errors being made on both higher and lower sides of the correct answer becomes balanced. By averaging the answers, the errors are mitigated due to the Law of the Large Numbers.

Currently, prevailing researches on the Wisdom of Crowds phenomenon primarily concentrate on the significance of diversity, decentralization, and independence. Nevertheless, fewer studies have delved into the aggregation methods that aid in consolidating a collective decision, although the process of consolidating a collective decision from a group of individuals with varying opinions and perspectives can be complex and challenging. Therefore, in this article, we will provide an overview of the aggregation methods utilized in previous studies on the wisdom of crowds. Understanding these methods is crucial in order to extract the collective wisdom from a group of individuals.

2 Methodology

Studying the wisdom of crowds spans across various fields, resulting in a substantial number of research papers. To address the problem statement, we utilize a traditional review method explained in [4].

The first step involves identifying the topic of the literature review. As mentioned above, our aim is to delve into the commonly employed aggregation methods in recent studies on the wisdom of crowds, as well as the reasons behind their preferred usage. In the second step, to find relevant papers, we employ Google Scholar as our search engine, utilizing keywords like "wisdom of crowd," "collective intelligence," and "aggregation method." To streamline our search process, we limit our exploration to approximately 10 result pages per keyword. We focus on retaining papers published in 2015 or later. For papers published prior to 2015, we establish a filtering criterion of 10 citations. Subsequently, in order to analyze and synthesize the literature, we proceed with an initial review of the collected articles to gain an understanding of their content. This involves reading the abstracts, introductions, and conclusions of the papers. After completing the initial overview, we revisit the articles and conduct a more systematic and critical review of their content, employing a structured approach proposed by Hendry and Farley [5].

3 Basic Notions

In this section, we briefly present basic notions related to the representation of the measure of central tendency and collective decision.

Types of Measure of Central Tendency

For comparison the condensation of data set, it is necessary to summarize the data set in a single value. Such a value usually somewhere in the center and represent the entire data set and hence it is called measure of central tendency or averages.

Definition 1. A value obtained by dividing the sum of all the observations by the number of observations is called *arithmetic mean* or simply as the *mean* [6].

Let $x_1, x_2, \dots, x_n$ be the data values, then we have

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

where $\bar{x}$ is a symbol representing the *mean* of the x_i values.

If the data are grouped, with f_i occurrences of the value x_i for $I = 1, 2, \dots, n$, then their mean is given by

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \tag{2}$$

where the numerator is the sum of all of the x_i values and the denominator is the total number of values.

Definition 2. A value obtained by the n^{th} root of the product of "*n*" positive values is called *geometric mean* [6].
Let $x_1, x_2, \dots, x_n$ be the data values, then the geometric mean (GM) is defined as:

$$\mathrm{GM} = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} \tag{3}$$

This can also be written as

$$\begin{aligned} \text{Log GM} &= \frac{1}{n}\log(x_1 \times x_2 \times \dots \times x_n) \\ &= \frac{1}{n}(\log x_1 + \log x_2 + \dots + \log x_n) = \frac{\sum \log x_i}{n} \end{aligned}$$

Therefore,

$$\mathrm{GM} = \text{Antilog}\,\frac{\sum \log x_i}{n} \tag{4}$$

where $n = f_1 + f_2 + \dots + f_n$

 ISBN: 978-604-80-8083-9

It is also represented as:

$$\mathrm{GM} = \sqrt[n]{\prod_{i=1}^{n} x_i} \tag{5}$$

For any grouped data, GM can be written as

$$\mathrm{GM} = \mathrm{Antilog}\,\frac{\sum f \log x_i}{n} \tag{6}$$

Definition 3. Let $x_1, \ldots, x_n \in R$ be a *sample* of estimates. The *collective decision* colD(x) is the aggregation of these estimates by an aggregation function, which is a statistical estimator of the *true value* θ [7].

$$\mathrm{colD}(x) = \bar{x} \tag{7}$$

For estimating a continuous value, the *cost function* is defined as a function of the difference of the *collective decision colD(x)* and the *true value* θ [7].

$$\text{cost function} = \mathrm{colD}(x) - \theta \tag{8}$$

Definition 4. The phenomenon of crowd wisdom can be conceptualized by *the diversity prediction theorem* as follows [7].

$$\mathrm{colErr}(x, \theta) = MSE(x, \theta) - Var(x) \tag{9}$$

Where $\mathrm{colErr} = (\bar{x} - \theta)^2$ is the *collective error*,

$\mathrm{MSE}(x, \theta) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \theta)^2$ is the *mean squared error*

and $\mathrm{Var}(x) = \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - x_i)^2$ is the *variance.*

4 Aggregation methods

In this section, we briefly provide an overview of the aggregation methods utilized in previous studies on the wisdom of crowds. There are various methods used for aggregating judgments from multiple individuals, which include the arithmetic average, geometric average, weighted average, median and mode average.

4.1 Arithmetic average

The fundamental approach was initially introduced by Francis Galton [8], who was the first to describe the Wisdom of the Crowd from a scientific perspective. If the decision-maker has limited knowledge about the domain, utilizing this basic average is a practical and reasonable approach. Implementing the arithmetic average assumes that all members of the crowd are assigned equal importance. Therefore, it appears that this average method has become a common aggregation method in various research studies [9], [10], [11], [12]. In addition, several studies have used arithmetic average as a standard for comparation with other methods of aggregation [13], [9],

[14]. In [14], the authors reported that the arithmetic mean and the median lead to very similar aggregates for participants' estimates and that neither is clearly superior to the other. The geometric mean turns out to be farther away from BBV in almost all cases, but only by a very narrow margin (Fig. 1).

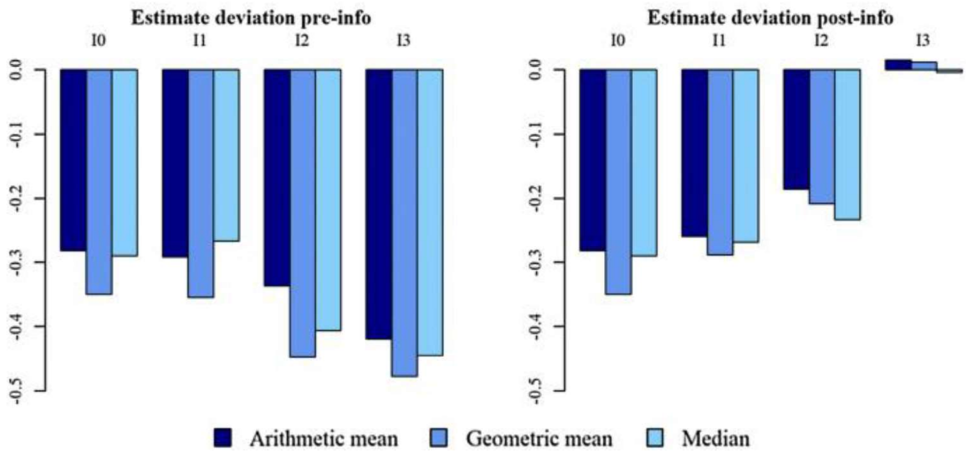

Fig. 1. Arithmetic and geometric mean and median log estimate deviation in units of BBV by jar, period and information level [14]

The benefit of the arithmetic average is that it can help reduce errors in predictions when the predicted values closely bracket the true values [15] and it easy to used [16]. On the other hand, one drawback of using the arithmetic average could be its impracticality in real-life scenarios, such as when multiple doctors need to evaluate a patient and make a prediction about their health over the course of a year [17]. When it comes to prediction polls, the arithmetic average tends to be underconfident, while the mode (most frequently occurring value) is often overconfident, as it assigns equal probabilities to each option in the poll [13], [18]. In addition, as stated in the paper by Lorenz [10], the arithmetic mean is not a suitable measure due to the right-skewed distribution of estimated values caused by the social influence effect. According to Atanasov [13], aggregating prediction polls and comparing them against prediction markets did not yield the best results. This is because prediction markets take into account updated predictions, individual skills, and correct for over- and under-confidence, whereas the arithmetic average used in prediction polls does not.

4.2 Geometric average

The geometric mean is a statistical measure that is used to determine the central tendency of a set of values. It is calculated by taking the nth root of the product of n values. The geometric mean has been used in various fields, including finance and ecology, to describe the average growth rate of a quantity over time. In recent years, researchers have started to investigate the potential use of the geometric mean in the context of the wisdom of crowds. In his article [10], Lorenz reported that the geometric mean is more appropriate for his data than the unweighted arithmetic average because the estimates of his type of questions are not normally distributed, but right-skewed (Fig. 2).

 ISBN: 978-604-80-8083-9

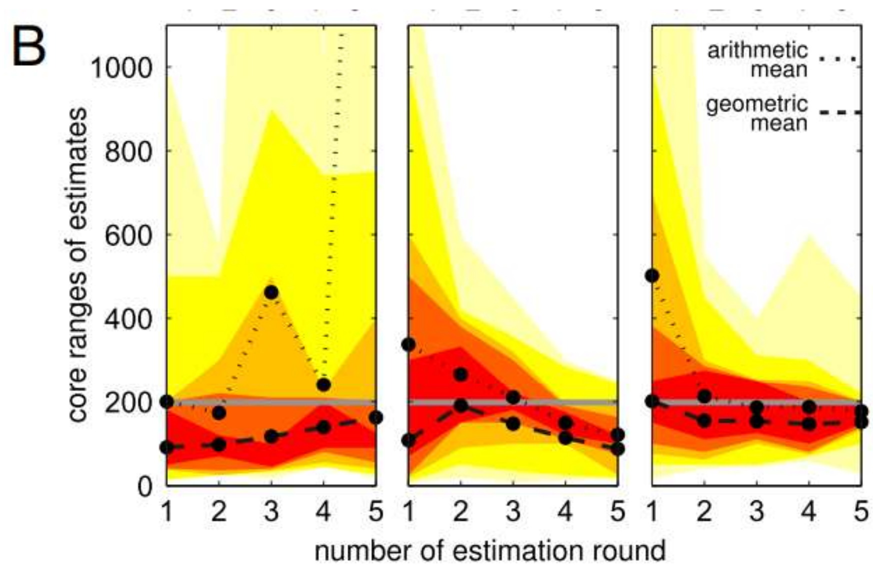

Fig. 2. Representation of the same data in aggregated form. The arithmetic mean is represented by a dotted line and the geometric mean by a dashed one [10]

Several studies involved conducting experiments to compare various aggregation methods, such as arithmetic and geometric means, and median, in different contexts of setting with asymmetric information. The study concluded that although the geometric average performed reasonably well, it was not the most effective approach [14].

In the context of studies on the role of social influence on the wisdom of crowds, which is a main direction of research, the geometric mean is often used to analyze the accuracy of group decisions. In their research [19], the authors reported that in cases where the solution space is wide and estimates will exhibit high variance and a wide range of positive values, the geometric mean is a more accurate measure of the wisdom of the crowd, as it captures the central tendency of the population better than the arithmetic mean.

4.3 The Median

Several studies have indeed examined the use of the median as a measure of central tendency in different contexts related to crowd predictions for comparing with other aggregation methods. Becker [9] used a prediction market to study the accuracy of crowd predictions for political events and found that the median forecast outperformed other measures of central tendency such as the mean or mode. Similarly, Hueffer [20] used a prediction market to forecast the spread of a disease and found that the median forecast was a better predictor of the actual outcome compared to other measures. Palan et al. [14] examined the use of the median as a measure of central tendency in survey responses and found that it improved the accuracy of the aggregated responses.

ISBN: 978-604-80-8083-9

An advantage of median is that it is not sensitive to outliers. This can be especially relevant in cases where respondents answer without any relevant knowledge or if they do not give much thought to their judgment [8].

4.4 Weighted aggregation

Weighted aggregation is a technique used in prediction that involves assigning a weight or importance factor to each individual's prediction. The weight is determined based on several factors such as their credibility, expertise, or past performance. In the study described in [21], the authors assigned weights based on self-assessment, knowledge, and hit rate in their research. On the other hand, Nguyen VD et al. in [22] assigned weights based on the individual's reputation. Some studies used different methods such as the accuracy of recent judgments [23], or summing probability predictions [24].

Weighting the crowd produced both positive and negative outcomes. When relying on the confidence of individual members, the crowd did not exhibit wisdom, as their personal biases led to overconfidence [12]. However, von der Gracht et al. [21] found that the individuals were not overconfident and exhibited more wisdom compared to individuals aggregated by their past performance. Despite the favorable outcomes, they did not discover any added benefits in weighting the crowd since equally weighting the responses resulted in superior outcomes. In their study presented in [25], the authors demonstrated that by removing underperforming individuals from the group and only considering those who made a positive contribution (model CWM), the overall performance of the crowd (which is highlighted in bold in Fig 3) improved significantly. This approach was found to be more effective than relying solely on past performance for weighting.

Model	Judges included	Mean positive contributors	$\bar{S}$ of models across all events				
			Min	Median	Mean	Max	SD
CWM	420	220	39.93	**91.90**	**88.26**	99.56	12.06
Contribution	420	220	39.52	89.55	86.46	99.50	11.82
UWM	420	—	41.58	87.45	83.73	98.25	11.51
ULinOp	1,233	—	42.81	87.64	83.62	98.67	11.76
xBWM	420	220	9.46	89.16	80.07	99.49	20.92
BWM	420	—	25.31	82.84	77.35	97.93	17.65

Fig. 3. Performance of the Models Compared (in Terms of Their Scores) [25]

Overall, the method of assigning weights to individuals depends on various factors such as the type of data being analyzed and the objectives of the research.

4.5 Mode

The mode can be a useful measure of central tendency when dealing with categorical data, such as words or other non-numerical values. For example, it can be used to determine which football player or team performs best [26], or which political party will win the election [27]. In cases where the data is categorical, it may not be appropriate to calculate a mean or median as these measures are typically used for numerical data. Instead, the mode can be used to identify the most common category or word in the dataset.

The mode works well as an aggregation method when individual predictions consist of one word and the crowd performs well [26], [27]. However, when it comes to ranking a larger number of items, other methods may be more suitable, such as using a scoring system or conducting pairwise comparisons. In their article [28], the authors reported that their Bayesian version of a Thurstonian model, which aggregates orderings across individuals, is more effectively (Rank = 87.0) than the mode (Rank=68.2) and other aggregation techniques (Fig. 4).

	Kemeny-Young			Thurstonian Model			Borda Counts			Greedy Count			Mode		
Problem	C	τ	Rank	C	τ	Rank	C	τ	Rank	C	τ	Rank	C	τ	Rank
books	0	4	96	0	6	88	0	7	82	0	7	82	0	12	40
city population europe	0	11	81	0	11	81	0	11	81	0	13	69	0	17	42
city population us	0	10	87	0	11	79	0	12	67	0	9	90	0	16	45
city population world	0	18	59	0	16	73	0	15	77	0	16	73	0	19	44
country landmass	0	7	76	0	5	95	0	5	95	0	5	95	0	7	76
country population	0	11	82	0	11	82	0	11	82	0	13	67	0	15	53
hardness	0	11	91	0	11	91	0	11	91	0	18	31	0	15	46
holidays	0	5	77	0	4	78	0	4	78	0	4	78	1	0	100
movies releasedate	0	2	95	0	2	95	0	2	95	0	2	95	0	2	95
oscar bestmovies	0	3	97	0	4	90	0	3	97	0	5	90	0	3	97
oscar movies	0	2	96	0	1	100	0	2	96	0	3	88	0	2	96
presidents	0	1	94	0	2	87	0	3	79	0	1	94	1	0	100
rivers	0	11	91	0	12	86	0	11	91	0	13	77	0	16	42
states westeast	0	1	97	0	2	88	0	3	78	0	1	97	0	1	97
superbowl	0	10	96	0	12	88	0	10	96	0	15	71	0	19	40
ten ammendments	0	2	97	0	4	95	0	5	90	0	4	95	0	4	95
ten commandments	0	11	82	0	11	82	0	12	74	0	12	74	0	17	51
AVERAGE	0.00	7.03	87.9	0.00	7.35	87.0	0.00	7.47	85.3	0.00	8.29	80.3	0.12	9.67	68.2

Fig. 4. Performance of the four heuristic models and the Thurstonian model [25]

4.6 Other aggregation methods

As data becomes more complex and diverse, simple aggregation methods like taking the mode or mean of responses may not be suitable or may lead to biased or inaccurate predictions. In such cases, more sophisticated aggregation methods may be needed to handle the complexity of the data and make more accurate predictions. However, these methods are more context-specific and not as commonly employed [29].

5 Conclusions and Future Works

This article presents a summary of the various aggregation methods used in the field of Wisdom of Crowds. Despite the growing number of studies involving large and complex data, the conventional aggregation methods, including arithmetic, geometric, weighted aggregation, and mode, remain prevalent in the Wisdom of Crowds domain.

Overall, the choice of aggregation method depends on the nature of the data and the goals of the study.

Our plan for future research is to investigate how social influence and aggregation methods relate to each other, with the aim of identifying the optimal range of applications for these methods. Social influence refers to the impact that one person's actions or opinions can have on the behavior or beliefs of others. In contrast, aggregation methods involve combining individual opinions or preferences to arrive at a group decision or consensus. Therefore, understanding the strengths and weaknesses of different methods in different situations could help identify the most appropriate approach to use for a given problem or decision-making process.

References

1. Surowiecki J. The Wisdom of the Crowds. New York City. United States: Anchor Books, a division of Random House; 2005.
2. Hong L, Page SE. Some microfoundations of collective wisdom. Collective wisdom. 2012:56-71.
3. Page SE. Where diversity comes from and why it matters? European Journal of Social Psychology. 2014;44(4):267-79.
4. Cronin P, Ryan F, Coughlan M. Undertaking a literature review: a step-by-step approach. British Journal of Nursing. 2008;17(1):38-43.
5. Hendry C, Farley AH. Reviewing the literature: a guide for students. Nursing standard (Royal College of Nursing (Great Britain) : 1987). 1998;12 44:46-8.
6. OpenStax. Introductory Business Statistics 2023. Available from: https://stats.libretexts.org/Bookshelves/Applied_Statistics/Introductory_Business_Statistics_(OpenStax).
7. Lorenz J. On the Quantification of Crowd Wisdom, Optimal Crowd Size, and the Mean-Median Dilemma. 2021.
8. Galton F. The ballot-box. Nature. 1907;75(1952):509-10.
9. Becker J, Brackbill D, Centola D. Network dynamics of social influence in the wisdom of crowds. Proceedings of the national academy of sciences. 2017;114(26):E5070-E6.
10. Lorenz J, Rauhut H, Schweitzer F, Helbing D. How social influence can undermine the wisdom of crowd effect. Proceedings of the national academy of sciences. 2011;108(22):9020-5.
11. Navajas J, Niella T, Garbulsky G, Bahrami B, Sigman M. Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. Nature Human Behaviour. 2018;2(2):126-32.
12. Simmons JP, Nelson LD, Galak J, Frederick S. Intuitive biases in choice versus estimation: Implications for the wisdom of crowds. Journal of Consumer Research. 2011;38(1):1-15.
13. Atanasov P, Rescober P, Stone E, Swift SA, Servan-Schreiber E, Tetlock P, et al. Distilling the wisdom of crowds: Prediction markets vs. prediction polls. Management science. 2017;63(3):691-706.
14. Palan S, Huber J, Senninger L. Aggregation mechanisms for crowd predictions. Experimental economics. 2020;23:788-814.
15. Hogg R, McKean J, Craig A. Introduction to mathematical statistics. Upper Saddle Rover. NJ: Prentice Hall; 2004.

 ISBN: 978-604-80-8083-9

16. Ariely D, Tung Au W, Bender RH, Budescu DV, Dietz CB, Gu H, et al. The effects of averaging subjective probability estimates between and within judges. Journal of Experimental Psychology: Applied. 2000;6(2):130.
17. Kattan MW, O'Rourke C, Yu C, Chagin K. The wisdom of crowds of doctors: Their average predictions outperform their individual ones. Medical Decision Making. 2016;36(4):536-40.
18. Wagner C, Suh A, editors. The wisdom of crowds: impact of collective size and expertise transfer on collective performance. 2014 47th Hawaii International Conference on System Sciences; 2014: IEEE.
19. Mavrodiev P, Schweitzer F. The ambigous role of social influence on the wisdom of crowds: An analytic approach. Physica A: Statistical Mechanics and its Applications. 2021;567:125624.
20. Hueffer K, Fonseca MA, Leiserowitz A, Taylor KM. The wisdom of crowds: Predicting a weather and climate-related event. Judgment and Decision making. 2013;8(2):91-105.
21. von der Gracht HA, Hommel U, Prokesch T, Wohlenberg H. Testing weighting approaches for forecasting in a Group Wisdom Support System environment. Journal of Business Research. 2016;69(10):4081-94.
22. Nguyen VD, Truong HB. Intelligent Collectives: Impact of Independence on Collective Performance. Cybernetics and Systems. 2022:1-16.
23. Mannes AE, Soll JB, Larrick RP. The wisdom of select crowds. Journal of personality and social psychology. 2014;107(2):276.
24. Murr AE. "Wisdom of crowds"? A decentralised election forecasting model that uses citizens' local expectations. Electoral Studies. 2011;30(4):771-83.
25. Budescu DV, Chen E. Identifying expertise to extract the wisdom of crowds. Management science. 2015;61(2):267-80.
26. O'Leary DE. Crowd performance in prediction of the World Cup 2014. European Journal of Operational Research. 2017;260(2):715-24.
27. Murr AE. The wisdom of crowds: What do citizens forecast for the 2015 British General Election? Electoral Studies. 2016;41:283-8.
28. Miller B, Hemmer P, Steyvers M, Lee MD, editors. The wisdom of crowds in rank ordering problems. 9th International conference on cognitive modeling; 2009: Citeseer.
29. Yi SKM, Steyvers M, Lee MD, Dry MJ. The wisdom of the crowd in combinatorial problems. Cognitive science. 2012;36(3):452-70.