Harvesting Wisdom on Social Media for Business Decision Making

Ji Yu
Massey University
j.yu3@massey.ac.nz

Nazim Taskin
Bogaziçi University
nazim.taskin@boun.edu.tr

David J. Pauleen
Massey University
d.pauleen@massey.ac.nz

Hamed Jafarzadeh
Massey University
h.jafarzadeh@massey.ac.nz

Abstract

The proliferation of social media provides significant opportunities for organizations to obtain wisdom of the crowds (WOC)-type data for decision making. However, critical challenges associated with collecting such data exist. For example, the openness of social media tends to increase the possibility of social influence, which may diminish group diversity, one of the conditions of WOC. In this research-in-progress paper, a new social media data analytics framework is proposed. It is equipped with well-designed mechanisms (e.g., using different discussion processes to overcome social influence issues and boost social learning) to generate data and employs state-of-the-art big data technologies, e.g., Amazon EMR, for data processing and storage. Design science research methodology is used to develop the framework. This paper contributes to the WOC and social media adoption literature by providing a practical approach for organizations to effectively generate WOC-type data from social media to support their decision making.

1. Introduction

Wisdom of the crowds (WOC) refers to the phenomenon that the aggregation of information in a group often generates solutions outperforming any solution by any individual member of the group [1]. WOC is appealing because it maximizes the amount of information available, reduces the potential impact of extreme or abnormal sources relying on incorrect or unreliable information, and increases the credibility and validity of the aggregation process by making it more inclusive and representative [2]. Currently, businesses are actively using WOC to support their work [3]. For example, in the UK, companies leverage WOC to acquire ideas for the improvement of their products [3].

Social media has become deeply embedded in people's daily lives. People rely on it for many needs, ranging from checking breaking news to connecting with friends [4]. Due to its proliferation, the content on social media, e.g., posts and reviews, has significantly contributed to the generation of big data [5, 6], which refers to "data sets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze." [7] (p. 143). The emergence of social media big data has led to a new growing trend of data analytics [6]. Organizations are increasingly using such analytics to accommodate this trend for achieving business goals such as increasing sales and enhancing customer satisfaction [8].

WOC-type data can also be generated through social media big data analytics. However, there are critical challenges associated with creating such data. For example, the openness of social media tends to increase the possibility of social influence, which may diminish group diversity, one of the conditions of WOC. In addition, although many studies have investigated WOC and social media big data analytics, practical guidance for organizations on how to leverage social media big data analytics to generate WOC for their decision making is scarce [9].

Therefore, the objective of this paper is to conceptualize a new social media data analytics framework to help organizations effectively generate, collect, and analyze WOC-type data on social media to support their decision making. The paper discusses the approaches to foster WOC (e.g., based on the issue at hand, using different discussion processes to overcome social influence issues and boost social learning). It also reviews the technologies in social media big data analytics that can be used for WOC generation (such as big data processing tools and services, e.g., Amazon EMR and Apache Spark, and data analytics methods, e.g., social network analysis and sentiment analysis). Through employing the appropriate approaches and technologies, the proposed framework offers an end-to-end solution for organizations to generate WOC-type data on social media for their decision making. Currently, the research is in progress.

The rest of the paper is structured as follows: the next section introduces the research background. This is followed by a discussion of research methodology

URI: https://hdl.handle.net/10125/80003
978-0-9981331-5-7

for developing the proposed new social media data analytics framework. The paper closes with conclusions, limitations, and future work.

2. Research background

2.1. Wisdom of the crowds

Wisdom is one of the most essential concepts of this study underpinning WOC. According to the Balance Theory [10], wisdom is the application of intelligence, creativity, and knowledge oriented toward a common good through the balance between intrapersonal, interpersonal, and extrapersonal interests.

WOC was introduced by Francis Galton in 1907 [11]. James Surowiecki used this concept in his book of the same name, in which he summarized the conditions of WOC: 1) diversity of group members who bring their private information, 2) ability of the members to think independently, 3) decentralization to leverage local and specific knowledge, and 4) the appropriate aggregation method to convert private opinions into a collective judgment [12].

Of the first two conditions, diversity helps because it brings opinions that would otherwise be missed and can reduce or remove some destructive components from group decisions [12]. Independence is beneficial since it excludes communication and information sharing processes or limits the effects of such processes [13]. Decentralization means that the decisional procedure has to be decentralized [12]. In a decentralized system, individuals may possess valuable information about many different aspects of the environment [14].

Diversity and independence are correlated with each other as people's opinions are more likely to converge (diversity suffers) when they exchange information (become dependent). The relevant concept here is social influence. One study [15] shows that social influence can cause people to lose independence, trigger the convergence of individual judgments, and considerably decrease group diversity without improving its accuracy.

However, social influence is not always detrimental. Studies demonstrate the ambiguous effect of social influence on WOC [16]. For instance, researchers identified scenarios in which WOC can be enhanced by increasing social influence when the initial collective error is high, and the initial average judgment is under the correct value. They also observed that WOC is decreased with the increase of social influence when the initial collective error is already low [16].

The reason that social influence is beneficial to the collective accuracy may be that participants can modify their opinions after learning from others, i.e., social learning, so the individual level precision increases [17]. However, this increased accuracy may not be able to offset the loss of the group-level diversity [17]. Hence, in order for social influence to strengthen WOC, participants should receive precise perceptions from others, and they should exchange opinions rather than merely follow others' decisions [17]. Unfortunately, research demonstrates that the crowd tends to focus more on exchanging members' judgments without elaborated discussions on the reasons behind such judgments [18].

Another related concept along with social influence and social learning is shared task experience. Study indicates when the shared task experience grows, the WOC performance increases [19]. These results are in line with previous research that a group's capability can be enhanced by members repeatedly working together [20, 21], but are contradictory to the view that the decreased diversity will have negative effects on WOC [22], suggesting that as the crowd gains more shared experience through working together, they can collectively adapt to the tasks more effectively [18].

Overall, for fostering diversity, independence, and decentralization, two aspects need to be considered: 1) group composition, e.g., the group size and members, and 2) processes.

For group size, although current technologies allow for gathering a large group of people, studies show that in many cases a small group of people can generate results outperforming a larger group (e.g., [23-25]). Müller-Trede et al. [24] illustrate that with the increase of group members the marginal benefits of additional judgment quickly diminish, and the optimal group size is about 10-15 people. This finding is consistent with the classic group aggregation research [26], which indicates that 8-12 group members can generate optimum performance levels. The costs of larger groups are usually higher than those of smaller groups [27], but this is more applicable for the judgment that requires some kind of expertise of group members [25]. Research shows that pooling the decisions from a small group of professional fingerprint experts generates solutions outperforming any solution by any individual member, while the decisions from a small group of novices produce even worse results [23].

For group members, obviously not all members are equal [28]. Some may be experts in the problem area, and some may potentially have more influence on others. Usually, people are more likely to be affected by influential members [28]. If such influencers

provide more accurate opinions, their followers may give more accurate answers as well. Conversely, if the influencers' opinions are less accurate, the collective judgment may be moved to the wrong direction [28]. Hence, seeking experts, who may offer more precise opinions, and weighting these opinions more strongly, can be helpful to leverage the advantages as well as combat the issues of the unequal member phenomenon [29]. Studies have investigated the weighted approaches. For example, in the case of picking stocks, Hill and Ready-Campbell [30] weighted participants' estimations based on the accuracy of their previous predictions, and they show that the weighted approach outperforms un-weighted one.

Regarding social demographic diversity, e.g., gender and religion, a study [31] suggests that they do not demonstrate significant effects on WOC. Socially diverse groups may produce similar opinions as socially homogeneous groups [31].

In terms of the WOC generation process, different situations may require different processes. When the decisions involve a large group of people (i.e., 30 or more people [32]), one open discussion stage may be enough as this method is helpful to encourage social learning; in particular the non-experts can learn from the experts, and the collective judgment is likely to reflect the opinions of higher weighted people, whose opinions tend to be more accurate [33]. When the decisions require a small group of experts (i.e., 29 or fewer with ideal size being 10-15 people [24]), e.g., doctors discussing the symptoms of a patient, aggregating the private opinions of individual members will be beneficial [33]. In this case, the WOC process can include three stages. First, before reviewing others' opinions, participants should express their own judgment. This step ensures that participants bring their diverse and independent opinions to the decision, so the social influence issues may be overcome. The second stage is to review each other's opinions and discuss the reasons behind the opinions. This approach allows participants to share their complementary information and build their judgment based on each other's knowledge, leading to boosting social learning [18, 34]. The third stage is to make their final judgment.

Finally, effective aggregation approaches need to be employed to access collective knowledge and make reliable decisions [14]. In general, studies show that the majority vote is the most common and easiest implemented method; it also demonstrates the greatest improvement in performance [23, 35]. Some researchers illustrate that under some conditions, averaging may be better than voting. For example, studies, e.g., [27, 28, 36], show that in truth-tracking tasks, averaging may generate more accurate results than voting. It is because the majority rule may make people concentrate too much on achieving agreement so that valuable minority judgments may be overlooked [36], whereas averaging considers more individual information and effectively neutralizes the over- and underestimations [17].

When voting and averaging prove suboptimal, people can resort to other approaches, such as wisdom of the resistant and choosing. For example, in their paper, Soll and Larrick [37] illustrate that if the crowd differs considerably in expertise, the optimal weights may be relatively extreme, and averaging might not be better than choosing or other methods.

2.2. Social media and WOC

The widespread use of social media provides a new channel for organizations to recruit diverse participants, collect rich, vast, and connected data, and conduct complex analysis for their decision making [38]. The links between the conditions of WOC and social media are illustrated in Table 1.

Table 1. Links between the conditions of WOC and social media

Conditions of WOC	Social Media
Diversity and independence	The proliferation of social media provides an opportunity to access a source sufficiently large to contain independent and diverse information, but social influence issues may still exist on social media [18, 39]. Hence, to leverage social media to acquire diverse and independent opinions, appropriate WOC generation approaches need to be employed.
Decentralization	Social media is not location specific. As long as people have an Internet connection, they can access social media. Therefore, individuals who possess local information can express their opinions through social media.
Aggregation	Social media big data analytics can be used to analyze and aggregate users' inputs.

Although the proliferation of social media provides an opportunity to access a source sufficiently large to contain independent and diverse information, its information sharing and communication facilitating nature makes purely independent judgments unlikely [18, 39]. Usually, social media users interact with each other directly. These interactions can cause social influence issues, as they may be convinced by others, and thereby produce interdependent and less diverse opinions [18, 40]. On the other hand, the social learning can also be promoted on social media. It can improve WOC as individuals share complementary information and revise their opinions based on each other's knowledge and experience [20, 34]. In particular, if the shared task experience increases in the crowd, it will strengthen WOC performance [18].

In general, the approaches to fostering WOC discussed in section 2.1 can be applied to the social media context. For example, depending on the issue at hand, organizations can assemble different kinds of groups, and use a weighted method on users' opinions.

2.3. Social media big data analytics

The popularity of social media has significantly increased the rate of data generation, resulting in big data [6], which is commonly defined by the "5Vs": 1) Volume is the magnitude of the data. Big data is much larger than normal datasets [41]; 2) Variety refers to the structural heterogeneity of datasets, i.e., structured, semi-structured, and unstructured datasets [41]; 3) Velocity refers to the speed of data generation [42]; 4) Veracity is about the accuracy of the data [42]; and 5) Value is how useful the data is in decision making [42].

Organizations are increasingly analyzing social media data to extract meaningful insights from the massive data to drive their decision making [9]. Social media big data analytics is the process of examining the big data on social media to uncover patterns and correlations, and other useful information [9]. It usually involves four stages. First, the data is collected from various social media sources and stored in big data storage, such as the Hadoop Distributed File System (HDFS) [6]. Then, as social media data may contain irrelevant and inconsistent information, it needs to be preprocessed to clean the data [6]. Third, in the processing and analysis stage, various technologies, such as Apache Spark, which is a unified analytics engine for large dataset processing, can be used to transform the data into meaningful insights [6, 43]. Finally, the data analysis results can be visualized for users to easily gain insights [6].

There are many useful social media big data analytics methods. This section presents the important ones in terms of generating WOC-type data for organizations in their decision making.

2.3.1. Semantic analysis is the interpretation of texts through natural language processing, which is concerned with exploring how computers can be used to understand and manipulate natural language text or speech for various tasks [44, 45].

Regarding analyzing participants' diversity, apart from common approaches, such as identifying people's occupation, expertise, and education level, semantic analysis techniques can also be applied. For example, Bhatt et al. [46] quantified the diversity of a group through the semantic analysis of the members' Twitter interactions. They measured the distance between group members by applying Word2Vec, a word embedding technique. It presents individual members within a high dimensional semantic vector space so that the diversity can be calculated according to the distance among members in the space [46]. The farther the members are apart, the more diverse they are [46]. To investigate the performance of the different groups, they identified the 500 most diverse (MD) groups and 500 least diverse (LD) groups based on their diversity scores. They also assembled 500 groups with randomly (R) selected members. They then asked each group to vote for the best captain for a fantasy team and evaluated which groups made better choices. The results show that MD groups outperform LD and R groups [46], demonstrating the usefulness of the diversity method.

2.3.2. Social network analysis. Another useful technology for identifying diversity is social network analysis, which is a process of analyzing the structure of a social network in order to explain the relationships of the network members and how the network operates [47]. In a social network, nodes are the individual members, and ties between them are their relationships. The ties can be divided into strong ties (i.e., people's connections with their family or ethnic group) and weak ties (i.e., people's connections outside their family or ethnic group) [48]. One study [49] indicates that strong ties make information become more influential within the group, which may lead to less diversity. On the other hand, weak ties are more likely to link heterogeneous members than strong ties are, resulting in the diversity of participants [50]. Hence, understanding the social ties among group members can assist in measuring the group diversity. If the group presents more weak tie characteristics, it tends to be more diverse. Social network analysis can be implemented by various techniques, such as counting the number of edges a node possesses and computing eigenvectors for identifying key nodes in a network [47].

2.3.3. Keywords extraction. When analyzing participants' discussions on social media, it is common to start with the keyword, which refers to "a word that succinctly and accurately describes the subject, or an aspect of the subject, discussed in a document" [51] (p. 341). The appropriate keywords may significantly improve the information retrieval efficiency and help users to quickly determine whether the information meets their requirements [52]. Hence, the quick and accurate extraction of keywords is crucial for social media data analytics [53].

In the social media context, because of the continuously increasing nature of the data, automatic keywords extraction is preferred [54]. It can be implemented through machine learning (ML) methods [55]. ML is a system with autonomous capability of acquisition and integration of knowledge learnt from experience and analytical observation [56]. ML methods can be categorized into four groups: 1) Supervised learning refers to using training datasets to learn the patterns between input and output data, and then predicting the output of test datasets by applying the identified patterns [57]; 2) Unsupervised learning allows the model itself to discover the patterns from the unlabeled datasets [57]; 3)Semi-supervised learning is a combination of supervised and unsupervised methods [57]; and 4) Reinforcement learning involves observing and taking actions to study which action can generate the most positive outcomes [58].

Usually, supervised or unsupervised methods are used in keywords extraction [55]. The main issue of supervised methods is that the training datasets required by the methods are often difficult to acquire, and they are usually domain-specific, so retraining may be necessary if the domain is changed [59]. Currently, a majority of keywords extraction methods are unsupervised, which can be divided into statistic-based, latent semantic analysis, and graph-based methods [55]. For example, Term Frequency-Inverse Document Frequency (TF-IDF) is a typical statistic-based keywords extraction method.

2.3.4. Sentiment analysis. To investigate participants' viewpoints, sentiment analysis, which belongs to the semantic analysis family, can be conducted. It is defined as a method to extract and understand people's opinions, e.g., positive, negative, and neutral, on text content. It is a critical task to gain meaningful insights for various purposes, such as obtaining product feedback from customers [60].

From a task-oriented perspective, sentiment analysis includes four features: polarity classification, beyond polarity, subjective and objective identification, and aspect-based analysis [61]. The most common sentiment analysis is polarity classification that identifies whether the opinion of a particular target is positive, negative, or neutral [61].

2.3.5. Topic modeling. To analyze users' discussions on social media, another very valuable text analytics technique is topic modeling. It is a collective term for a family of computational algorithms aiming to help users to uncover the themes behind the unlabeled documents [62].

Topic modeling relies only on a few assumptions of the text data, so it has been applied to a wide range of sources [63]. Some most commonly used topic modeling methods are Latent Semantic Analysis (LSA) [64], Latent Dirichlet Allocation (LDA) [65] and Non-negative Matrix Factorization (NMF) [66].

Currently, LDA has become one of the most popular methods in the topic modeling family [67]. The key concept behind LDA is an imaginary generative process assuming that a document consists of a discrete group of topics and each topic includes a discrete distribution of words [63]. Each document displays the topics in different proportions, from 0% (i.e., a document does not present a topic at all) to 100% (i.e., a document fully focuses on a topic) [63]. Using LDA, users can quickly figure out the themes in the social media discussion.

2.4. Organizational decision making

According to Ackoff [68], organizational decisions can be categorized into operational, tactical and strategic ones. Operational decisions, such as daily task assignments and production planning, are primarily routine and well-defined [68]. Usually, managers should make decisions quickly based on current information and data. Tactical decisions, such as quarterly operating schedules and inventory control, are mid-term decisions and often related to organizational efficiency and effectiveness [68]. While some relevant data can be gathered for this kind of decision, it may be not worth spending too much time and cost to collect considerable amount of data to support them [69]. They tend to rely on managers' judgments [69]. Strategic decisions, such as product innovation, and merger and acquisition, are long-term decisions, and they have profound influences on organizational goals [68]. They tend to involve a large amount of data and information. Considering the importance and necessity, strategic decisions are more likely to employ WOC to support judgments.

3. Design science research methodology

A design science research methodology (DSRM) approach is applied in this study as it intends to

propose a new social media data analytics framework to solve an identified problem. Table 2 summarizes the stages of the DSRM approach.

Table 2. DSRM stages [70]

Stages	Description
Problem identification and motivation	Define the research problem and justify the value of the solution.
Define objectives	Infer the objective of the solution, which is to develop a new social media data analytics framework to help organizations generate, collect, and analyze WOC-type data for their decision making.
Design and development	Implement the framework to address the explicated problems and meet the defined requirements.
Demonstration	A case will be conducted to prove the feasibility of the framework.
Evaluation	Employ a case study method to determine whether the framework is effective in solving the explicated problems.
Communication	This stage will be performed through publications (e.g., journal papers and magazine articles).

The first stage is problem identification and motivation. As mentioned in the above sections, WOC can be a very powerful approach to reach a near-to-optimal solution in a group. The mass uptake of social media and its significant advantages offer an important opportunity for organizations to generate WOC-type data. However, there are still various issues, such as social influence, associated with the data generation. Practical instructions guiding organizations to leverage social media big data analytics to create WOC-type data for their decision making are also rare [9]. Hence, this paper intends to fill these gaps. Its objective is to conceptualize a new social media data analytics framework to help organizations effectively generate, collect, and analyze WOC-type data on social media to support their decision making. In the third stage, the framework will be designed and developed. After that, in the demonstration stage, a case will be performed to prove the feasibility of the framework. In the evaluation phase, a multiple case study will be employed to determine whether the framework is effective in solving the explicated problems. The communication stage will be conducted through publications (e.g., conference and journal papers, and magazine articles) to reach academic and practitioner communities. The DSRM process is iterative, so the demonstration, evaluation, and communication stages can provide critical feedback to enhance the design and development [71]. Currently, the framework has been designed and is in the development phase. The next sections will discuss this phase and the following demonstration phase in detail.

3.1. Design and development

3.1.1. WOC generation strategy. The framework architecture is shown in Figure 1. First, organizations should consider the issue at hand. As discussed in section 2.1, if it is a decision requiring participants to offer opinions based on their expertise, perhaps assembling a small group of people with a specific knowledge set is a good choice [23-25]. For example, organizations may invite their business partners and suppliers to discuss decisions related to their supply chains. If the decision requires the perspectives from a large group of people, recruiting such a group may be necessary [20, 34]; for instance, organizations making decisions associated with their product enhancement or new product development.

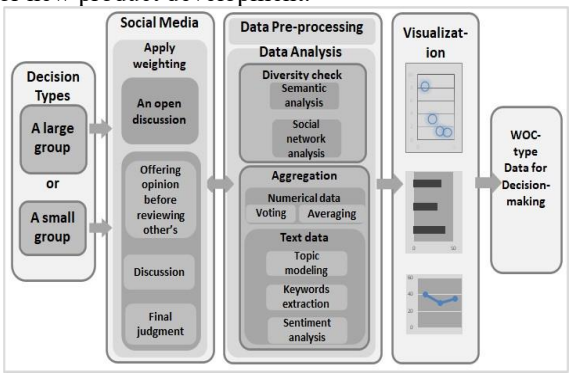

Figure 1. Framework architecture

Regarding the group diversity, for a small group, this can be ensured by verifying each member's background, such as their education level, expertise, and professional roles. For a large group, novel approaches, such as identifying diversity through semantic analysis or social network analysis mentioned in sections 2.3.1 and 2.3.2, can be used.

Organizations can also adopt a weighted approach, e.g., finding experts through investigating participants' previous performance and assigning a higher weight to the people whose opinions tend to be more accurate [29].

For the WOC generation process, as mentioned in section 2.1, different processes can be employed based on the decision. If the decision involves a large group of people, one open discussion stage may be enough,

whereas if the decision requires a small group of experts, the WOC process can include three stages (i.e., participants express their own opinions, review each other's opinions and discuss the reasons behind the opinions, and then make their final judgments).

3.1.2. Data collection and analysis. After deciding the WOC generation strategy, organizations may establish the environment for data collection and analysis. This may include the following steps.

First, organizations should set up the big data infrastructure, which may include a group of machines to store and process data. It can be an internal or external environment. Small- and Medium-sized Enterprises (SMEs) may lack the capability to build an internal big data environment. Hence, they can employ some public cloud resources, such as Amazon Web Services (AWS), which provide online cloud computing platforms and APIs to customers [72]. In this paper, AWS is used as an example. Figure 2 presents the data analysis process and the technologies used in it.

To handle big data, a single machine is usually inadequate, so a cluster, which refers to a group of coordinated machines, needs to be set up to distribute the workload to different machines [73].

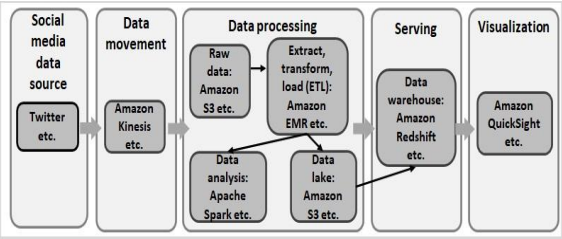

Figure 2. Data analysis process [74]

The technologies such as Amazon EMR, which is a big data platform to quickly and effectively process huge amounts of data, can be employed to set up a cluster, and Apache Spark can be used to process data in the cluster environment. A cluster (see Figure 3 [75]) typically includes one master node and several worker nodes - a node is an individual machine in the cluster [73]. It can be managed by using tools such as Apache YARN, which is formed by two types of daemons: a ResourceManager running on the master node and NodeManager(s) running on the worker node(s) [76]. The ResourceManager has two major components: Scheduler and ApplicationsManager [76]. Spark jobs can be submitted on YARN. In the cluster mode, when a Spark application is launched, the Spark driver runs inside an ApplicationMaster, which is created by the ApplicationsManager for each Spark application to negotiate resources from the ResourceManager [75]. After getting the resources, Spark executors will run tasks on worker node(s) [75].

Organizations should also have a data warehouse that is a central repository to store the organized data from various sources, e.g., databases, for example, using Amazon Redshift, which is a quick, fully managed, petabyte-scale data warehouse [74].

Decision makers can access the data through business intelligence tools, such as Amazon QuickSight, which is a cloud-powered business intelligence service making it easy for users to build visualizations, to make more informed decisions [74].

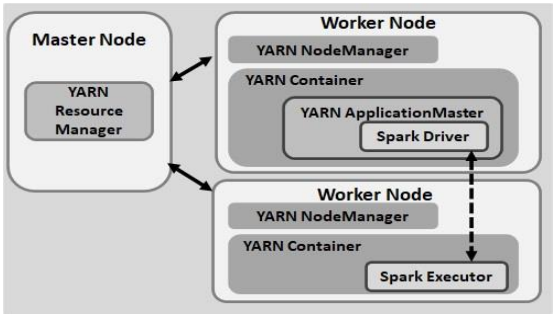

Figure 3. A Spark cluster

Having the big data environment established, organizations can then initiate appropriate groups and start the discussions. The data can be collected and analyzed in real time by using the APIs from social media providers. For example, if they use Twitter, they can collect tweets through the Twitter Streaming API, and a service such as Amazon Kinesis, which is a fully managed service for streaming data real-time processing, can be employed to load and store the raw data in data storage, e.g., Amazon Simple Storage Service (Amazon S3), for further purposes [72].

Next, the Spark jobs can be run to access the data in Amazon S3, and then preprocess the data and conduct analysis.

Due to the unstructured nature of social media data, pre-processing needs to be conducted to structurally transform the data. It includes data cleaning for incorrect information, tokenization, stop words removal, and lemmatization [77].

Then, the methods such as semantic analysis and social network analysis can be performed to identify the diversity of group members, and the group may be adjusted if necessary. After that, based on the decision type, organizations can choose different aggregation methods. For example, for decisions requiring a prediction or estimation number, majority voting or averaging may be employed. For sophisticated problems which require more interactions and deliberations of participants, text analytics methods, e.g., keywords extraction, sentiment analysis, and topic modeling, may be used to aggregate and gain the insights from people's discussions.

Finally, the data analysis results can be visualized on a dashboard, such as on Amazon QuickSight [78], for decision makers.

3.2. Demonstration

In the demonstration phase, a single, instrumental case study will be performed to prove the feasibility of the framework and gain a broader appreciation [79]. The case is designed to be conducted in a medium- or large-sized software development company, which plans to develop a new product. Product innovation is a key strategy for companies to create a sustained competitive advantage [80]. Software development company is selected, as this type of company is more likely to have already built up the data analytics infrastructure and have the data analysis capability.

To adopt the framework, the first step is to check the appropriate big data infrastructure. If it has not been set up, the researchers of this study will help the company to build the environment and apply the framework. Second, as the strategic goal is product innovation, it may be better to invite a large number of stakeholders, such as customers, employees, and business partners, to an open discussion to tap their collective insights. The company's social media channel, such as their own website or their Facebook pages, can be used for this purpose. A weighted approach may be applied to weight participants' opinions [29]. For example, software architects and key customers may be assigned higher weights due to their expertise or importance to the company. After generating the ideas, corresponding data collection, processing and analysis will be conducted, and then based on the company's needs, customized reports can be generated and visualized for senior leadership team to make decisions. Finally, for the data collection of this case study, interviews and surveys will be performed to gather participants' feedback about the framework. The feedback will be analyzed by the researchers of this study and corresponding actions will be taken to improve the framework. The experience gained in this case will be employed to design a multiple case study in evaluation phase.

4. Conclusions

In this study, a new social media data analytics framework was introduced to help organizations effectively generate, collect, and analyze WOC-type data for their decision making. The paper illustrated the research problem, motivation, and objective, and summarized the significant literature. The methodology used in this study was also discussed. Currently, the architecture of the framework has been proposed. In future work, the framework will be implemented and tested on several selected organizations to evaluate its effectiveness at solving the explicated problems.

The potential limitations of the framework include: 1) The capability of this framework depends on the actual participants organizations recruit (e.g., the number and diversity). If they cannot recruit the desired participants, the generated WOC might not be able to meet their needs; and 2) The framework is still in the development phase. It needs to be implemented and applied to organizations to prove its feasibility and usefulness.

5. References

[1] R. Baeza-Yates and D. Saez-Trumper, "Wisdom of the Crowd or Wisdom of a Few?: An Analysis of Users' Content Generation," in *Proceedings of the 26th ACM Conference on Hypertext & Social Media,* pp. 69-74, *2015.*

[2] D. V. Budescu and E. Chen, "Identifying Expertise to Extract the Wisdom of Crowds," *Management Science,* vol. 61, no. 2, pp. 267-280, 2015.

[3] M. Hosseini, J. Moore, M. Almaliki, A. Shahri, K. Phalp, and R. Ali, "Wisdom of the crowd within enterprises: Practices and challenges," *Computer Networks,* vol. 90, pp. 121-132, 2015.

[4] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi, and S. Nerur, "Advances in Social Media Research: Past, Present and Future," *Information Systems Frontiers,* vol. 20, no. 3, pp. 531-558, 2018.

[5] K. Lyu and H. Kim, "Sentiment analysis using word polarity of social media," *Wireless Personal Communications,* vol. 89, no. 3, pp. 941-958, 2016.

[6] N. A. Ghani, S. Hamid, I. A. T. Hashem, and E. Ahmed, "Social media big data analytics: A survey," *Computers in Human Behavior,* vol. 101, pp. 417-428, 2019.

[7] S. Yin and O. Kaynak, "Big data for modern industry: challenges and trends," in *Proceedings of the IEEE,* vol. 103, no. 2, pp. 143-146, 2015.

[8] W. He, S. Zha, and L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry," *International journal of information management,* vol. 33, no. 3, pp. 464-472, 2013.

[9] W. He, F.-K. Wang, and V. Akula, "Managing extracted knowledge from big social media data for business decision making," *Journal of Knowledge Management,* vol. 21, no. 2, pp. 275-294, 2017.

[10] R. J. Sternberg, "Older but not wiser? The relationship between age and wisdom," *Ageing International,* vol. 30, no. 1, pp. 5-26, 2005.

[11] F. Galton, "Vox Populi," *Nature,* vol. 75, no. 1949, pp. 450-451, 1907.

[12] J. Surowiecki, *The Wisdom of Crowds*. Anchor Books, 2005.

[13]P. Mavrodiev, C. J. Tessone, and F. Schweitzer, "Effects of Social Influence on the Wisdom of Crowds," *arXiv:1204.3463*, 2012.

[14] A. Laan, G. Madirolas, and G. G. de Polavieja, "Rescuing Collective Wisdom when the Average Group Opinion Is Wrong," *Frontiers in Robotics and AI,* vol. 4,

no. 56, 2017.
[15] J. Lorenz, H. Rauhut, F. Schweitzer, and D. Helbing, "How social influence can undermine the wisdom of crowd effect,"in *Proceedings of the National Academy of Sciences,* vol. 108, no. 22, pp. 9020-9025, 2011.
[16] P. Mavrodiev and F. Schweitzer, "The ambigous role of social influence on the wisdom of crowds: An analytic approach," *Physica A: Statistical Mechanics and its Applications,* vol. 567, p. 125624, 2021.
[17] C. Ganser and M. Keuschnigg, "Social influence strengthens crowd wisdom under voting," *Advances in Complex Systems,* vol. 21, no. 06n07, p. 1850013, 2018.
[18] B. Yan, L. Jian, R. Ren, J. Fulk, E. Sidnam-Mauch, and P. Monge, "The Paradox of Interaction: Communication Network Centralization, Shared Task Experience, and the Wisdom of Crowds in Online Crowdsourcing Communities," *Communication Research,* vol. 48, no. 6, pp. 796-818, 2021.
[19] C. Riedl and V. P. Seidel, "Learning from mixed signals in online innovation communities," *Organization Science,* vol. 29, no. 6, pp. 1010-1032, 2018.
[20] A. Almaatouq, A. Noriega-Campero, A. Alotaibi, P. Krafft, M. Moussaid, and A. Pentland, "The Wisdom of the Network: How Adaptive Networks Promote Collective Intelligence," *arXiv preprint arXiv:1805.04766,* 2018.
[21] L. Argote, B. L. Aven, and J. Kush, "The effects of communication networks and turnover on transactive memory and group performance," *Organization Science,* vol. 29, no. 2, pp. 191-206, 2018.
[22] H. Hong, Q. Ye, Q. Du, G. A. Wang, and W. Fan, "Crowd characteristics and crowd wisdom: Evidence from an online investment community," *Journal of the Association for Information Science and Technology,* vol. 71, no. 4, pp. 423-435, 2020.
[23] J. M. Tangen, K. M. Kent, and R. A. Searston, "Collective intelligence in fingerprint analysis," *Cognitive research: principles and implications,* vol. 5, no. 23, pp. 1-7, 2020.
[24] J. Müller-Trede, S. Choshen-Hillel, M. Barneron, and I. Yaniv, "The wisdom of crowds in matters of taste," *Management Science,* vol. 64, no. 4, pp. 1779-1803, 2018.
[25] M. Galesic, D. Barkoczi, and K. Katsikopoulos, "Smaller crowds outperform larger crowds and individuals in realistic task conditions," *Decision,* vol. 5, no. 1, pp. 1-15, 2018.
[26] R. M. Hogarth, "A note on aggregating opinions," *Organizational behavior and human performance,* vol. 21, no. 1, pp. 40-46, 1978.
[27] E. Libby and L. Glass, "The calculus of committee composition," *PLoS One,* vol. 5, no. 9, p. e12642, 2010.
[28] Z. Da and X. Huang, "Harnessing the wisdom of crowds," *Management Science,* vol. 66, no. 5, pp. 1847-1867, 2020.
[29] A. B. Kao *et al.*, "Counteracting estimation bias and social influence to improve the wisdom of crowds," *Journal of The Royal Society Interface,* vol. 15, no. 141, p. 20180130, 2018.
[30] S. Hill and N. Ready-Campbell, "Expert stock picker: the wisdom of (experts in) crowds," *International Journal of Electronic Commerce,* vol. 15, no. 3, pp. 73-102, 2011.
[31] S. De Oliveira and R. E. Nisbett, "Demographically diverse crowds are typically not much wiser than homogeneous crowds," in *Proceedings of the National Academy of Sciences,* vol. 115, no. 9, pp. 2066-2071, 2018.
[32] L. Kreeger, *The large group: Dynamics and therapy*. Routledge, 2019.
[33] U. Hertz, M. Romand-Monnier, K. Kyriakopoulou, and B. Bahrami, "Social influence protects collective decision making from equality bias," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 42, no. 2, pp. 164-172, 2016.
[34] J. Becker, D. Brackbill, and D. Centola, "Network dynamics of social influence in the wisdom of crowds," in *Proceedings of the national academy of sciences,* vol. 114, no. 26, pp. E5070-E5076, 2017.
[35] A. Vercammen, Y. Ji, and M. Burgman, "The collective intelligence of random small crowds: A partial replication of Kosinski et al.(2012)," *Judgment and Decision Making,* vol. 14, no. 1, pp. 91-98, 2019.
[36] J. Lorenz, H. Rauhut, and B. Kittel, "Majoritarian democracy undermines truth-finding in deliberative committees," *Research & Politics,* vol. 2, no. 2, pp. 1-10, 2015.
[37] J. B. Soll and R. P. Larrick, "Strategies for revising judgment: How (and how well) people use others' opinions," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 35, no. 3, pp. 780-805, 2009.
[38] L. Sloan and A. Quan-Haase, *The SAGE handbook of social media research methods*. SAGE Reference, 2017.
[39] F. Franch, "(Wisdom of the Crowds)2: 2010 UK Election Prediction with Social Media," *Journal of Information Technology & Politics*, vol. 10, no. 1, pp. 57-71, 2013.
[40] L. Muchnik, S. Aral, and S. J. Taylor, "Social influence bias: A randomized experiment," *Science,* vol. 341, no. 6146, pp. 647-651, 2013.
[41] J. Mneney and J.-P. Van Belle, "Big data capabilities and readiness of South African retail organisations," in *6th International Conference-Cloud System and Big Data Engineering*: IEEE, pp. 279-286, 2016.
[42] Y. Gahi, M. Guennoun, and H. T. Mouftah, "Big data analytics: Security and privacy challenges," in *IEEE Symposium on Computers and Communication*, pp. 952-957, 2016.
[43] Apache. "Apache spark," Available: https://spark.apache.org/
[44] G. G. Chowdhury, "Natural language processing," *Annual review of information science and technology,* vol. 37, no. 1, pp. 51-89, 2003.
[45] S. A. Salloum, R. Khan, and K. Shaalan, "A survey of semantic analysis approaches," in *Joint European-US Workshop on Applications of Invariance in Computer Vision*: Springer, pp. 61-70, 2020.
[46] S. Bhatt, B. Minnery, S. Nadella, B. Bullemer, V. Shalin, and A. Sheth, "Enhancing crowd wisdom using measures of diversity computed from social media data," in *Proceedings of the International Conference on Web*

Intelligence, pp. 907-913, 2017.
[47] I. Lee, "Social media analytics for enterprises: Typology, methods, and processes," *Business Horizons,* vol. 61, no. 2, pp. 199-210, 2018.
[48] M. Woolcock and D. Narayan, "Social capital: Implications for development theory, research, and policy," *The world bank research observer,* vol. 15, no. 2, pp. 225-249, 2000.
[49] G. Weimann, "The strength of weak conversational ties in the flow of information and influence," *Social networks,* vol. 5, no. 3, pp. 245-267, 1983.
[50] M. S. Granovetter, "The strength of weak ties," in *Social networks*: Elsevier, pp. 347-367, 1977.
[51] J. Feather and P. Sturges, *International encyclopedia of information and library science*. Routledge, 2003.
[52] L. Zhang, X.-P. Li, F.-B. Zhang, and B. Hu, "Research on Keyword Extraction and Sentiment Orientation Analysis of Educational Texts," *Journal of Computers,* vol. 28, no. 6, pp. 301-313, 2017.
[53] S. Pan, Z. Li, and J. Dai, "An improved TextRank keywords extraction algorithm," in *Proceedings of the ACM Turing Celebration Conference*, pp. 1-7, 2019.
[54] H. Shah, M. U. Khan, and P. Fränti, "H-rank: a keywords extraction method from web pages using POS tags," in *IEEE 17th International Conference on Industrial Informatics*, vol. 1: IEEE, pp. 264-269, 2019.
[55] D. Zhao, N. Du, Z. Chang, and Y. Li, "Keyword extraction for social media short text," in *14th Web Information Systems and Applications Conference*: IEEE, pp. 251-256, 2017.
[56] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
[57] A. Dey, "Machine learning algorithms: a review," *International Journal of Computer Science and Information Technologies,* vol. 7, no. 3, pp. 1174-1179, 2016.
[58] M. Pecht, "Prognostics and health management of electronics," in *Encyclopedia of structural health monitoring,* John Wiley & Sons, 2009.
[59] H. Ma, S. Wang, M. Li, and N. Li, "Enhancing graph-based keywords extraction with node association," in *International Conference on Knowledge Science, Engineering and Management*: Springer, pp. 497-510, 2019.
[60] T. Shivaprasad and J. Shetty, "Sentiment analysis of product reviews: a review," in *2017 International Conference on Inventive Communication and Computational Technologies*: IEEE, pp. 298-301, 2017.
[61] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowledge and Information Systems,* vol. 60, no. 2, pp. 617-663, 2019.
[62] S. K. Ray, A. Ahmad, and C. A. Kumar, "Review and Implementation of Topic Modeling in Hindi," *Applied Artificial Intelligence,* vol. 33, no. 11, pp. 979-1007, 2019.
[63] S. Debortoli, O. Müller, I. Junglas, and J. vom Brocke, "Text mining for information systems researchers: An annotated topic modeling tutorial," *Communications of the Association for Information Systems,* vol. 39, no. 7, pp. 110-135, 2016.
[64] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse processes,* vol. 25, no. 2-3, pp. 259-284, 1998.
[65] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research,* vol. 3, pp. 993-1022, 2003.
[66] J. Nassour, D. Leykin, M. Elhadad, and O. Cohen, "Computational Text Analysis of A Scientific Resilience Management Corpus: Environmental Insights and Implications," *Journal of Environmental Informatics,* vol. 36, no. 1, pp. 24-32, 2020.
[67] D. Chehal, P. Gupta, and P. Gulati, "Implementation and comparison of topic modeling techniques based on user reviews in e-commerce recommendations," *Journal of Ambient Intelligence and Humanized Computing,* vol. 12, pp. 5055–5070, 2021
[68] R. L. Ackoff, "Strategy," *Systems practice,* vol. 3, no. 6, pp. 521-524, 1990.
[69] S. Gressel, D. J. Pauleen, and N. Taskin, *Management Decision-making, Big Data and Analytics*. SAGE, 2020.
[70] K. Peffers, T. Tuunanen, M.A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45-77, 2007.
[71] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly,* vol. 28, no. 1, pp. 75-105, 2004.
[72] J. Varia and S. Mathew, "Overview of amazon web services,"Available: http://cabibbo.dia.uniroma3.it/asw-2014-2015/altrui/AWS_Overview.pdf
[73] S. Sujitha and S. Jaganathan, "Aggrandizing Hadoop in terms of node heterogeneity & data locality," in *International Conference on Smart Structures & Systems*: IEEE, pp. 145-151, 2013.
[74] Amazon, "Analytics on AWS," Available: https://aws.amazon.com/big-data/datalakes-and-analytics/
[75] C. Lin, "How to Build a Real-Time Twitter Analysis Application Using Big Data Tools," Available: https://towardsdatascience.com/how-to-build-a-real-time-twitter-analysis-using-big-data-tools-dd2240946e64
[76] Apache, "Apache Hadoop YARN," Available: https://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html
[77] S. Liu and I. Lee, "A Hybrid Sentiment Analysis Framework for Large Email Data," in *10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp. 324-330, 2015.
[78] R. Nadipalli, *Effective Business Intelligence with QuickSight*. Packt Publishing Ltd, 2017.
[79] R. E. Stake, *The art of case study research*. Sage, 1995.
[80] E. Danneels, "The dynamics of product innovation and firm competences," *Strategic management journal,* vol. 23, no. 12, pp. 1095-1121, 2002.