An Information Fusion based Framework for Social Big Data

Ibrar Yaqoob, Umer Majeed, and Choong Seon Hong

Department of Computer Engineering, Kyung Hee University, Yongin-si, 17104, Korea.

{ibraryaqoob, umermajeed, cshong}@khu.ac.kr

Abstract

Social big data is a budding research area that is drawing the interest of IT communities. Social big data is derived from social networking sites such as Facebook, Twitter, and Instagram. The advent of modern communication technologies has enabled the extraction of information by synergistically integrating data from various sources of social media. This study is organized from the point of view of applying information fusion (IF) techniques to social big data. First, we explore the benefits of applying IF to social big data and highlight its current trends. Furthermore, we propose an IF-based framework, which is comprised three layers. The qualitative analysis results reveal that utilization of IF reduces uncertainty, which leads to mitigating risks in a proactive manner, and helps in decision-making.

1. Introduction:

Finding accurate information in social big data is becoming a challenge for public and private institutions. The emergence of new distributed technologies in social big data management requires accurate information [1]. The sources of information generated from social media are ever increasing and can be useful in predicting user trends [2]. Collecting data from various interrelated sources can provide complete data while enabling a better understanding of the observed trend by reducing ambiguity. Information fusion (IF) consists of combining information into a new set of information with the goal of reducing the uncertainty of the data. The fusion of diverse information sources is a suitable solution for the operational needs of social big data analytics, and IF may allow companies to achieve their analytic goals more efficiently.

Figure 1 shows the characteristics of social big data. A wide range of social media data makes data processing within a limited time span difficult [3]. Undoubtedly, some of the data used to meet the needs of organizations for decision-making purposes are obtained through social media [4]. When information originates from multiple sources, as in social media, IF can help produce superior results. After the fusion of data, applying analytical techniques can help to extract valuable and accurate information [5-6].

The contributions of this study are as follows: (1) We explore the benefits of applying IF to social big data and highlight its current trends. (2) We propose an IF-based framework which helps to mitigate risks in a proactive manner. These contributions are provided in the separate sections from 2-3. The study is summarized in section 4.
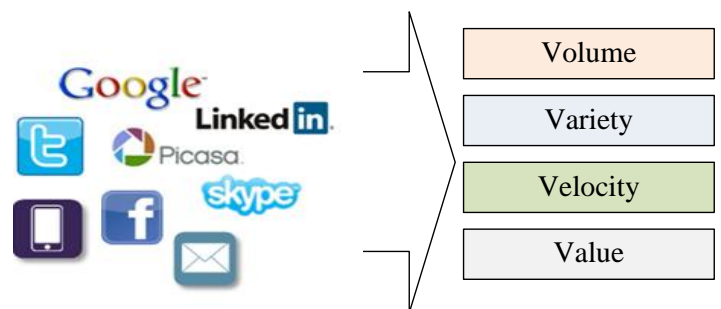

Figure 1: Characteristics of social big data

2. Motivation and Trends :

IF is necessary to be applied in social big data because of its following benefits:

- Fusion increases the accuracy of information generated from multiple sources by reducing the uncertainty of data. For instance, social

sensors such as Facebook and Twitter provide redundant information that can increase the strength of the system.

- The fusion of complementary information generated from multiple sources can provide more accurate information instead of single-source information.

- Confidence can be increased when multiple independent measurements are made on the same event dataset. Therefore, the result is more reliable.

- Multiple social sensors reduce the set of hypotheses about the target, hence, the ambiguity of the measured information can be reduced.

- The inherent redundancy of information from different sources can increase reliability.

The constant trend of social media applications demands efficient data management [7]. Moreover, the usability of social media applications such as Facebook, Twitter, LinkedIn, and Instagram has significantly increased in several domains including business, government, and social networking [8]. Table 1 compares the technologies used to handle social big data. Smart analysis of social media data can provide the following benefits:

- Security agencies can prevent fraud by analyzing a huge dataset of social media.

- Marketers can analyze the interests of the people in the business perspective.

- Various types of systems can be recommended by analyzing the content of users.

- Advance feasibility studies of businesses can be performed by analyzing the interest of people.

- Politicians can determine the needs of the people and their level of satisfaction with the government.

3. Proposed Framework:

Figure 2 illustrates a theoretical framework proposed for large-scale social media datasets, which is comprised three layers.

The first layer enables the users to share their data using social networking sites, whereas, the second layer is responsible for storing, processing, and analyzing the social big data in a real-time manner, as it consists of edge server and data center. Additionally, the second layer ensures the utilization of IF techniques, as all users' data are usually managed by multiple cloud service providers. The third layer is responsible for data visualization, which helps in decision-making. Adapting IF in social big data can provide many benefits, such as advanced marketing, fraud detection, social context-based recommendation systems, an advanced feasibility study for new businesses, and maximum decision accuracy.

Table 1: Comparison of the social big data processing technologies

|	Hadoop	Storm	Apache Spark	Apache Tez	Cloud Dataflow	Corona
Programming API	Java	Java, Scala, Python, Others	Java, Scala, Python	Java, Python, Others	Java, Python	Java, Python, Others
Platform	Yarn	Storm Cluster, Yarn	Yarn, Mesos, Standalone	Yarn, MapReduce	Google Cloud and BigQuery	Hadoop MaReduce
Latency	Seconds	Milliseconds	Seconds	N/A	N/A	Milliseconds
Resource Efficiency Usage	Low	High	High	Medium	High	High
Database Supports	HDFS	HDFS	HDFS, Cassandra, Hbase and S3	HDFS	Bigtable	N/A
Implemented In	Java	Clojure	Scala	Java	Java	Java
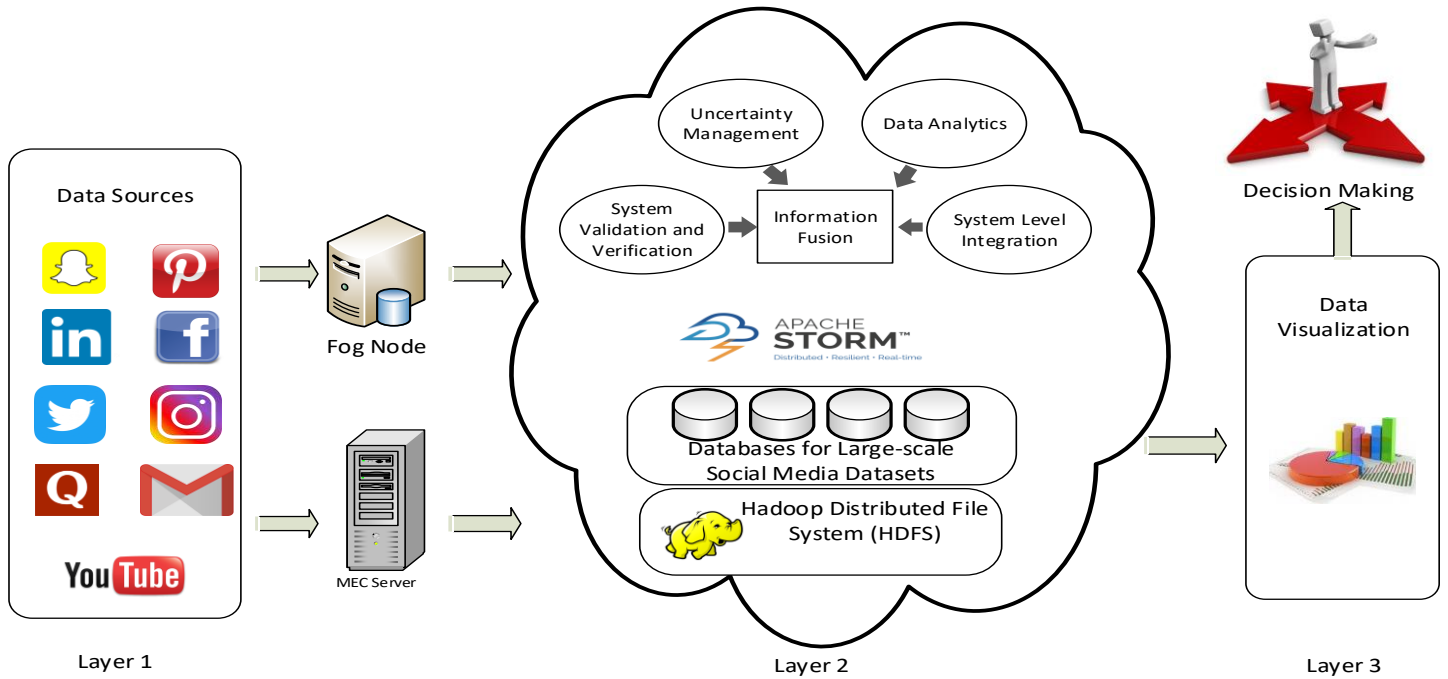
Figure 2: An illustration of the proposed framework

4. Summary:

During the past decade, a need for IF has emerged from social big data applications, giving an upswing to an emergent area of research that is still in its infancy.

In this paper, we initially explored the benefits of applying IF to social big data and highlighted its current trends through motivation. Then, we proposed a theoretical IF-based framework for social big data which helps to mitigate risks in a proactive manner. This study encourages researchers to apply IF to social big data because it is a very effective solution. However, promising results in the field do not conceal the challenges that still need to be addressed before any widespread deployment of IF technologies in social big data can be undertaken.

Acknowledgement:

[References]

[1] Bello-Orgaz, Gema, Jason J. Jung, and David Camacho. "Social big data: Recent achievements and new challenges." *Information Fusion* 28 (2016): 45-59.

[2] Akbar, Adnan, et al. "Real-Time Probabilistic Data Fusion for Large-Scale IoT Applications." *IEEE Access* 6 (2018): 10015-10027.

[3] Farasat, Alireza, et al. "Social network analysis with data fusion." *IEEE Transactions on Computational Social Systems* 3.2 (2016): 88-99.

[4] Chen, Siming, et al. "Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data." *IEEE transactions on visualization and computer graphics* 22.1 (2016): 270-279.

[5] Wang, Chih-Yu, Yan Chen, and KJ Ray Liu. "Game-Theoretic Cross Social Media Analytic: How Yelp Ratings Affect Deal Selection on Groupon?." *IEEE Transactions on Knowledge and Data Engineering* 30.5 (2018): 908-921.

[6] Jabbar, Sohail, et al. "A Methodology of Real-Time Data Fusion for Localized Big Data Analytics." *IEEE Access* 6 (2018): 24510-24520.

[7] Kim, Jooho, and Makarand Hastak. "Social network analysis: Characteristics of online social networks after a disaster." *International Journal of Information Management* 38.1 (2018): 86-96.

[8] Ramírez-Gallego, Sergio, et al. "Big data: tutorial and guidelines on information and process fusion for analytics algorithms with MapReduce." *Information Fusion* 42 (2018): 51-61.