Lecture Notes in Electrical Engineering 551

Jason C. Hung
Neil Y. Yen
Jia-Wei Chang *Editors*

# Frontier Computing

## Theory, Technologies and Applications (FC 2019)

Springer

Jason C. Hung · Neil Y. Yen · Jia-Wei Chang
Editors

# Frontier Computing

## Theory, Technologies and Applications (FC 2019)

Springer

*Editors*
Jason C. Hung
Department of Computer Science
and Information Engineering
National Taichung University of Science
and Technology
Taichung, Taiwan

Jia-Wei Chang
Department of Computer Science
and Information Engineering
National Taichung University of Science
and Technology
Taichung, Taiwan

Neil Y. Yen
School of Computer Science
and Engineering
The University of Aizu
Aizu-Wakamatsu, Japan

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-15-3249-8 ISBN 978-981-15-3250-4 (eBook)
https://doi.org/10.1007/978-981-15-3250-4



This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Measuring Public Opinion with Social Media Use in Local Government of Asian Cities

Shih-Nung Chen[1(✉)], Ridho Al-Hamdi[2], Yong-Kok Tan[1], Aulia Nur Kasiwi[2], and Achmad Nurmandi[2]

[1] Department of Information Communication, Asia University, Taichung City, Taiwan, R.O.C.
nung@asia.edu.tw
[2] Department of Government Affairs and Administration, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia

**Abstract.** Social media enables government to discover events in real time, and forecast public opinion. This study presents a system prototype for measuring public opinion from News channels, Bulletin Board Systems (BBS) and social networking sites, including Facebook. The proposed system aims to improve communication between government officials and ordinary citizens about service delivery. The proposed system applies event-driven simulation to accelerate the processing speed, and thus provides a better solution for measuring public opinion.

**Keywords:** Social media · Public opinion measuring · TF-IDF · Event-driven simulation · Distributed processing

## 1 Introduction

Research into the use of social media by the government in making a policy or a set of policies is important, because government traditionally consults people through formal face-to-face meetings. These formal meetings do not always represent all people, because they do not allow people to deliver their opinions frankly to the policy maker, while social media enables people to deliver their own aspirations, and to express personal emotions to the government directly without any pressure.

With the rapid growth of Internet users, the Internet has become the main channel for people to obtain and publish information, increasing the impact of network information on public opinion. The huge amount of information on the ever-changing Internet underline the importance of collecting and analyzing the information of network public opinion quickly; finding hotspots and sensitive topics; performing tracking analysis of topics and forecasting network public opinion warning; timely and effectively responses to the dynamic changes in the network of public opinion, and supporting and guaranteeing the correct guidance of public opinion [1].

The purpose of this study is to develop a smart automatic system for monitoring and analyzing the domestic and foreign public opinion on political issues through News, Bulletin Board Systems (BBS), Facebook and other social networking sites. This distributed public opinion measuring system processes the characteristics of public


J. C. Hung et al. (Eds.): FC 2019, LNEE 551, pp. 399–406, 2020.
https://doi.org/10.1007/978-981-15-3250-4_48

opinion data based on text extraction, English and Chinese word segmentation and semantic analysis, and text classification. The system adopts event-driven simulation to enhance the efficiency of the measuring process, and using Python to develop the system, thus improving user insight on people's support for policies, as well as satisfaction level on government's decision towards future policies through the proposed system.

## 2 Related Work

Governments pay attention to public opinion, and many government officials find the views of the public through the media. The rise in popularity of social media services, and increased ability of computers to process huge volumes of information at high speeds, has necessitated and enabled automatic content analysis of mass media in the social sciences.

Social media provides additional sources of politically relevant information to people in almost real time. The majority of the work using social media data applied to political performance predictions is information accounting from a given social network during a pre-defined period of time. The challenge of developing text classification systems for social media is dealing with accounting for the significant lexical variation resulting the informal nature of the text [2]. Social media is easy to use, and provides a cheap way of communication. These functionalities enable the use of social media as a collaborative opinion mining platform, and motivate the use of social media for opinion trend analysis [3–6].

### 2.1 Opinion Tracking and Trend Analysis

Social media data can be applied for many purposes, especially public opinion tracking. A good subjective information collection can be found on Forums, which amongst other topics, has political discussions covering the entire spectrum of politics related issues. Such data can be collected through systems that can automatically track public opinion enabling decision makers to make sense of the enormous body of opinions expressed on the social media [3].

Topic tracking technology is can track massive Internet information, according to a main topic, and can automatic screen and detect new topics for tracking. Trend analysis is defined as the analysis of changes over time. Opinion tracking and trend analysis technology are targeted to specific, usually focus events occurring within a certain period of time, such as hot news and sensitive events. The analysis involves the event itself and associations with other events, public opinion and social impact [7]. This study mainly concentrates on opinion trend analysis. The proposed system is intended to help government officials to see the most trending government service delivery related topics on SNSs.

### 2.2 Text to Quantization Model and TF-IDF Algorithm

The gist of text can usually be expressed in the keywords or index terms. The index terms are based on the main content of the text, and enable the reader to quickly overall understanding of the whole text [7].

In the study of text feature selection, an extraction algorithm constructs a weighting function, and independently evaluates characteristics of the concentration of each feature to obtain assessment points. The points are then sorted, and are used to construct a subset of features in reservation number. The number of features to be selected, and the evaluation function to use, depends on the precision and speed of text classification for decision [8, 9]. Text feature selection evaluation functions include term frequency, inverted document frequency [10], IG [11], x2 statistics [12], expected cross entropy, mutual information [13], weight of evidence text, and odds ratio [14]. The TF-IDF algorithm is a commonly utilized approach to calculate keyword weights, in order to describe the text characteristics accurately.

The TF-IDF formula has two main components, namely TF (Term Frequency) and IDF (Inverse Document Frequency). The principle of TF-IDF is that if a word appears in a document with high frequency and rarely appears in other documents, then it has very good distinguishing ability and high suitability for classification. A commonly used TF-IDF formula is as follows [8]:

$$w(t, d) = tf(t, d) * log\left(\frac{N}{N_t}\right) \quad (1)$$

The formula $tf(t, d)$ denotes the frequency of keyword $t$ in text $d$; $N$ represents the number of full text, and $N_t$ represents the number of the texts which have the word $t$.

## 3 Design and Implementation

Public opinion research is applying new methods, such as automatic data extraction, data mining, language processing, information intelligent discovery and statistics, to design and realize public opinion information discovery. This study presents a distributed public opinion measuring integrated platform in Chinese and English.

### 3.1 Overall Design of the System

The proposed system has the following steps: collect information from various News, BBS, Facebook, or other social networking sites; perform filtering and scheduling to survey and track topics; use distributed processing to improve efficiency; analyze huge amounts of public opinion data, and finally show the measuring results. Therefore, the system must consider the following issues.

1. Solve the problem of retrieving data from News, BBS, Facebook and other social networking sites.
2. Design opinion tracking.
3. Design a distributed processing mechanism.

4. Analyze huge amounts of public opinion data.
5. Design a public opinion measuring system.

### 3.2 System Component Function Design

Figure 1 shows the overall design and system architecture of the system, which is divided into five functional components:

1. The data acquisition component performs data acquisition and preprocessing.
2. The opinion tracking component performs hot topic and sensitive topic selection and tracking, URL filtering and scheduling.
3. The distributed processing component performs the rapid processing of data access.
4. The public opinion analysis component performs noise elimination, text extraction, English and Chinese segmentation and semantic analysis, then transforms the data to unified format data and performs classification.
5. The public opinion measuring component includes system configuration, and performs public opinion measuring.

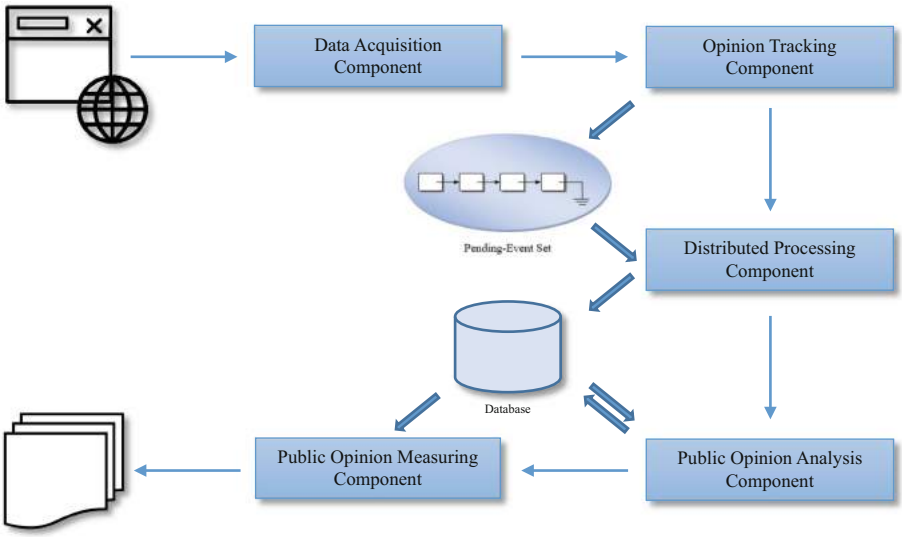


**Fig. 1.** System architecture

The main flow of the distributed processing component is as follows:

1. The URL is generated according to requirements.
2. The URL is inserted at the tail of the event queue to be crawled.
3. The distributed crawler extracts the URL from the first of the event queue.
4. The distributed crawler downloads the public opinion content.
5. The downloaded public opinion content is stored in the database.
6. The public opinion content is analyzed.

7. The distributed crawler extracts the URL to be crawled in the queue to download the work, if the queue URLs to be crawled is not empty, then steps 3–7 are repeated.
8. The application obtains the index data, supports the measuring function, and finally presents the measuring results to the user in form of public opinion briefs.

### 3.3 Discrete-Event Simulation

To perform distributed processing, the concept of discrete-event simulation is added to the design of the system [15, 16]. The proposed system is assumed to comprise many processes, each comprising a series of events. Distributed processing component picks up the first event of the pending-event set to simulate and calculate the simulation time and insert new event to pending-event set as necessary. The component continues to repeat the actions until no event exists in the pending-event set. The pending-event set is a data structure storing every scheduled event that has not yet occurred. The events are held in sequence according to the time of occurrence. The data structure of each event includes the event type and the time of occurrence, as shown in Fig. 2.

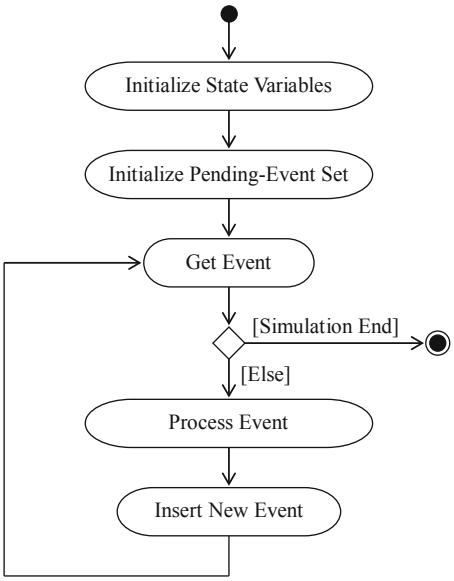


**Fig. 2.** Discrete-event simulation

### 3.4 Scheduling Synchronization and Load Balancing

The synchronization between distributed processing components is achieved using logical clock theory proposed by Lamport [17]. Logic clock is an abstract concept. A clock is a value assigned to a specific event, and the value denotes the time that the event will occur. Function $C$ denotes the whole time system. The total ordering relationship $\Rightarrow$ is defined as follows. If $a$ denotes an event of process $P_i$, and $b$ denotes an

event of process $P_j$, then when $a \Rightarrow b$, if and only if not (1) $C_i<a> < C_j<b>$, then (2) $C_i<a> = C_j<b>$ and $P_i < P_j$.

To use computing resources and balance system loading effectively [18–23], the system has default scheduling rules which can be adjusted by the user. The AppLeS (Application Level Scheduling) methodology [24] was used to determine the default scheduling rules. The individual steps followed by our scheduling rules are shown in Fig. 3 and detailed below.

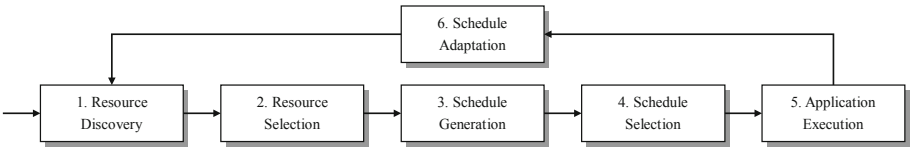


**Fig. 3.** Steps in the AppLeS methodology

## 4 Conclusion

Social media can easily provide and obtain information, and drive innovation in public service delivery and government operation. Social media amplifies the power of citizens' opinions on the Internet to affects public information and public services. This study presents a system prototype for measuring public opinion in Chinese and English from various News, BBS, Facebook, or other social networking site. The proposed system applies event-driven simulation for distributed processing to accelerate the processing speed, thus provides a better solution for measuring public opinion and using in local government of Asian cities.

**Acknowledgment.** The authors would like to thank the Asia University of the Republic of China, Taiwan, for financially supporting this research under Contract No. 106-ASIA-UMY-02.

## References

1. Shi, Z.Q., Chen, S.P.: The design and implementation of opinion extraction system based on distributed network. In: Proceedings of International Industrial Informatics and Computer Engineering Conference, pp. 1335–1338 (2015)
2. Saleiro, P., Amir, S., Silva, M., Soares, C.: POPmine: tracking political opinion on the web. In: Proceedings of IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pp. 1521–1526 (2015)
3. Mfenyana, S.I., Moroosi, N., Thinyane, M., Scott, S.M.: Development of a Facebook crawler for opinion trend monitoring and analysis purposes. Int. J. Comput. Appl. **79**(17), 32–39 (2013)
4. Srivastava, M.: Social media and its use by the government. J. Public Adm. Gov. **3**(2), 161–172 (2013)

5. Bassett, R., Chamberlain, T., Cunningham, S., Vidmar, G.: Data mining and social networking sites: protecting business infrastructure and beyond. J. Inf. Syst. **11**(1), 352–357 (2010)
6. Costa, E., Ferreira, R., Brito, P., Bittencourt, I.I., Holanda, O., Machado, A.: A framework for building web mining applications in the world of blogs: a case study in product sentiment analysis. Expert Syst. Appl. **39**(5), 4813–4834 (2012)
7. An, D.Z., Wu, G.L., Lu, J., Zhang, S.C., Li, Y.: Design and implementation of network public opinion monitoring and analysis system. In: Proceedings of the 1st EAI International Conference on Multimedia Technology and Enhanced Learning (2017)
8. Wang, W., Tang, Y.X.: Improvement and application of TF-IDF algorithm in text orientation analysis. In: Proceedings of International Conference on Advanced Material Science and Environmental Engineering, pp. 230–233 (2016)
9. Li, M.T., Luo, J.Y., Yin, M.J.: Combining with the meaning of the text of key weight calculation method. Comput. Appl. **32**(5), 1355–1358 (2012)
10. Lu, Y.C., Lu, M.Y.: In the vector space method word weighting function analysis and structure. Comput. Res. Dev. 1205–1210 (2002)
11. Mladenic, D., Crobelnik, M.: Feature selection for unbalanced class distribution and naive bayes. In: Proceedings of the 16th International Conference Machine Learning (1999)
12. Yang, Y., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: Proceedings of the 14th International Conference on Machine Learning (1997)
13. Yang, Y.M., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: Proceedings of the 14th International Conference on Machine Learning, pp. 412–420 (1997)
14. Kenneth, W.C., Patrick, H.: Word association norms, mutual information and lexicography. In: Proceedings of ACL 27, pp. 76–83 (1989)
15. Misra, J.: Distributed discrete-event simulation. ACM Comput. Surv. **18**(1), 39–65 (1986)
16. Chou, C.C.: Parallel simulation and its performance evaluation. Technical report 93-02, Ph. D. Dissertation, Department of Computer Science, The University of Iowa, Iowa City, USA (1993)
17. Lamport, L.: Time, clocks, and the ordering of events in a distributed system. Commun. ACM **21**(7), 558–565 (1978)
18. Kondo, D., Casanova, H., Wing, E., Berman, F.: Models and scheduling mechanisms for global computing applications. In: Proceedings of the International Parallel and Distributed Processing Symposium (2002)
19. Shao, G.: Adaptive scheduling of master/worker applications on distributed computational resources. Ph. D. thesis, University of California, San Diego (2001)
20. Takefusa, A., Matsuoka, S., Nakada, H., Aida, K., Nagashima, U.: Overview of a performance evaluation system for global computing scheduling algorithms. In: Proceedings of the Eighth IEEE International Symposium on High Performance Distributed Computing, pp. 97–104 (1999)
21. Faerman, M., Su, A., Wolski, R., Berman, F.: Adaptive performance prediction for distributed data-intensive applications. In: Proceedings of the IEEE/ACM SC 99 Conference, pp. 36–50 (1999)
22. Maheswaran, M., Ali, S., Siegel, H.J., Hensgen, D., Freund, R.F.: Dynamic matching and scheduling of a class of independent tasks onto heterogeneous computing systems. In: Proceedings of the Eighth Heterogeneous Computing Workshop, pp. 30–44 (1999)

23. Foster, I., Kesselman, C., Lee, C., Lindell, B., Nahrstedt, K., Roy, A.: A distributed resource management architecture that supports advance reservations and co-allocation. In: Proceedings of the International Workshop on Quality of Service, pp. 27–36 (1999)
24. Berman, F., Wolski, R., Casanova, H., Cirne, W., Dail, H., Faerman, M., Figueira, S., Hayes, J., Obertelli, G., Schopf, J., Shao, G., Smallen, S., Spring, N., Su, A., Zagorodnov, D.: Adaptive computing on the grid using AppLeS. IEEE Trans. Parallel Distrib. Syst. **14**(4), 369–382 (2003)