

Research Article

The Finding and Dynamic Detection of Opinion Leaders in Social Network

Baocheng Huang,¹ Guang Yu,¹ and Hamid Reza Karimi²

¹ School of Management, Harbin Institute of Technology, Harbin 150001, China

² Department of Engineering, Faculty of Engineering and Science, The University of Agder, 4898 Grimstad, Norway

Correspondence should be addressed to Baocheng Huang; huanghjlhrbin@gmail.com

Received 23 December 2013; Accepted 19 January 2014; Published 4 March 2014

Academic Editor: Xudong Zhao

Copyright © 2014 Baocheng Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is valuable for the real world to find the opinion leaders. Because different data sources usually have different characteristics, there does not exist a standard algorithm to find and detect the opinion leaders in different data sources. Every data source has its own structural characteristics, and also has its own detection algorithm to find the opinion leaders. Experimental results show the opinion leaders and their characteristics can be found among the comments from the Weibo social network of China, which is like Facebook or Twitter in USA.

1. Introduction

With further study, the definition of opinion leader expands. It involves not only the most influential person but also the most influential commentary. The finding and detection of opinion leader in social network have great commercial and political values. By identifying the most influential person, companies or governments can use this feature for selling or guiding public opinion, respectively. Additionally, detecting the most influential comments is also able to understand the source of public opinion formation process. By building multiple topic networks, this essay can detect opinion leaders with the algorithm of POLD (Positive Opinion Leader Detection). Some researchers had found that some ideas in control field such as data driven [1–6] and robust control [7–13] could be into the study of find and dynamic detection of opinion leaders in social network. However, this idea only stays on the think level. Therefore, this work will propose a Dynamic Opinion Rank algorithm to find the opinion leaders in the comments of Chinese news. By using the methodology, it can find the most influential comments from all the network comments and the most influential users form the entire user network.

2. Problem Formulation

A single theme network based Weibo news consists of three levels. Those levels are themes, comments, and users. There exist some relationship mappings between those different levels. For example, the mapping between themes and comments is 1 divided by n , while the mapping between comments and users is m divided by n . This study will analyze a single topic, build a single view based network and the mathematical model of users, and then find out the most influential comments and users. The structure of those three levels is shown in Figure 1.

As shown in Figure 1, there are three levels. “Layer 1” stands for all the themes of news, “Layer 2” denotes the single-topic network $G_{CN}(V, E)$, which is composed of comments, while “Layer 3” is the single-topic user network $G_{UN}(V, E)$. Based on $G_{CN}(V, E)$, it is possible to find the most influential comments cmt_3 . Then, by using the mapping between comments and users, it can find out U_3 from $G_{UN}(V, E)$.

Definition 1 (the most influential comments). For the comment sets $C = \{C_1, C_2, \dots, C_n\}$, $G_{CN}(V, E)$ is a single point of network theme. After sorting by using some algorithms, there

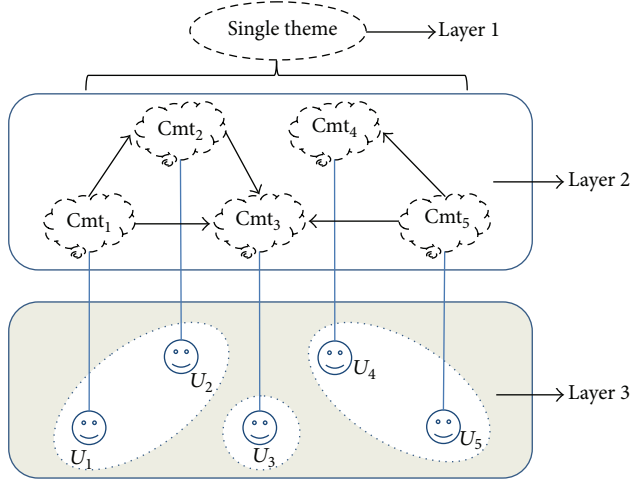


FIGURE 1: The model of single theme network model.

TABLE 1: Sentiment analysis of forum comments.

The contents of the comments	Emotional tendencies
Comment Number 1	Positive
Comment Number 2	Negative
Comment Number 3	Neutral

is sore $C_{i, \text{score}}$ for any comment $C_i \in C$. Based on the sort according to their scores, it can be assumed that $C_{1, \text{score}} > C_{2, \text{score}} > \dots > C_{m, \text{score}}$. Then, the comments with the highest score are defined as the most influential comments; these are also the opinion leaders of comments [14].

Definition 2 (the most influential user). As in Definition 1, for the user set $U = \{u_1, u_2, \dots, u_m\}$, each user $u_i \in U$ has its own score $u_{i, \text{score}}$. Sorting those scores, it follows that $u_{1, \text{score}} > u_{2, \text{score}} > \dots > u_{m, \text{score}}$. Then, the user with the highest score is the most influential one; it is defined as the opinion leader of users [15].

3. Attitude Stabilization Controller Design

3.1. Analysis of Emotions. The work of finding the positive and negative emotion links requires determining the propensity of emotions. Based on the HowNet dictionary, this work will firstly determine the emotional feelings of tendentious comments [16]. Comments are usually classified according to emotional bias: positive (1), negative (-1), and neutral (0); this is shown in Table 1.

According to the preceding definitions, any comments $C_i \in C$ ($1 \leq i \leq n$) can be divided into m statements, such as $\langle S_1, S_2, S_3, \dots, S_m \rangle$. Then, it applies ICTCLAS to split each statement S_j ($1 \leq j \leq m$) into one word, such as $\langle w_1, w_2, w_3, \dots, w_l \rangle$ [17], and it extracts emotion words from emotion dictionary and obtains the number of statistical statements S_j containing negative words, such as "No." Usually, emotional value of words is set as 1, -1, and

0. Finally, accumulating all the emotional words in S_j , it can obtain the emotional value of the statement. Using the parity negative word sentence to correct the statement of emotional tendencies, it yields a final statement $\{-1, 1, 0\}$, while the cumulative review of all statements yields final emotional tendencies $\{1, -1, 0\}$.

3.2. Modeling of Single-Topic Network. By using the explicit and implicit link algorithm, all of the link relationship of the set C is found. Based on the sentiment analysis methods, the algorithm proposed by this work to establish a single theme network is described as in Algorithm 1.

In Algorithm 1, the 1st to 7th line is to traverse the set C and to find out all of the link. The 8th is going to give a certain weight wt_{ij} to the positive and negative links; this weight wt_{ij} is given by

$$wt_{i,j} = \text{tag} \times \text{similarity}(C_i, C_j) \quad (-1 \leq wt_{i,j} \leq 1). \quad (1)$$

In this equation, the function $\text{similarity}(C_i, C_j)$ represents the contents of comments similarity between C_i and C_j and "tag" denotes the emotional consistency between comments. For any reply relationship $C_i \rightarrow C_j$, if the comment is consistent tendency, then this comment is viewed as a positive link; that is, $\text{tag} = 1$. Otherwise, the comment is a negative link and $\text{tag} = -1$. The weight is thus assigned according to the following equation:

$$wt_{i,j} = \begin{cases} 1 & c_j \text{ explicitly links to } c_i \text{ and tag} = 1 \\ -1 & c_j \text{ explicitly links to } c_i \text{ and tag} = -1 \\ \text{Sim}(c_i, c_j) & c_j \text{ implicitly links to } c_i \text{ and tag} = 1 \\ -\text{Sim}(c_i, c_j) & c_j \text{ implicitly links to } c_i \text{ and tag} = -1. \end{cases} \quad (2)$$

In (2), if the connection relationship between C_i and C_j is explicit, then the similarity is equal to 1. If it is an implicit link relationship, the similarity is between the texts. If the emotional tendency between C_i and C_j is consistent, then the weight $wt_{i,j}$ will not be changed, otherwise, its value will become opposite. The above construction procedure can be illustrated as in Figure 2.

In Figure 2(b), the structure of the set $C = \{C_1, C_2, C_3, C_4, C_5\}$ is explicitly evaluated. According to the chronological order release $C_1 < C_2 < C_3 < C_4 < C_5$, and the sequence corresponding to the floors floor₁, floor₂, floor₃, floor₄, and floor₅, the corresponding single topic views networks as shown in Figure 2(a). The serial number of the edge is ranked in ascending order. That represents the link discovery order.

4. Dynamic Detection of Opinion Leaders

Opinion leaders are the most influential comments or persons. This paper will present an approach to find out the most influential comments among a single point of network $G_{CN}(V, E)$ and build a user views the network $G_{UN}(V, E)$ to find out the most influential user.

```

Input: explicit links and implicit links in C, sentiment orientation  $O_i$  of every  $C_i \in C$ ;
Output:  $G_{CN}(V, E)$  //Comment Network of C;
Description:
(1) for each  $C_i \in C$ 
(2)   for each  $C_j = C_i \in C$ 
(3)     if ( $C_i$  link to  $C_j$ ) //the link includes explicit and implicit link
(4)     if  $C_i$  has the same sentiment orientation with  $C_j$ 
(5)        $C_i$  positive link to  $C_j$ ;
(6)     else
(7)        $C_i$  negative link to  $C_j$ ;
(8)   assign weight  $wt_{ij}$  for edge  $C_i \rightarrow C_j$ ;
    
```

ALGORITHM 1: The algorithm to build comment network.

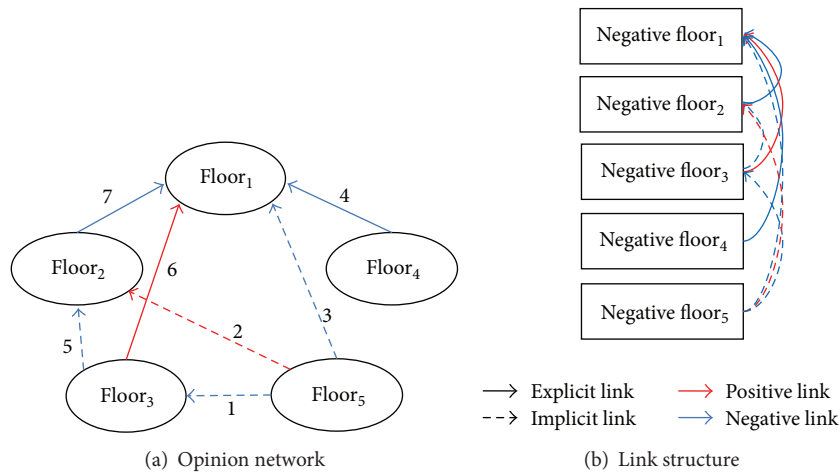


FIGURE 2: Single theme opinion network.

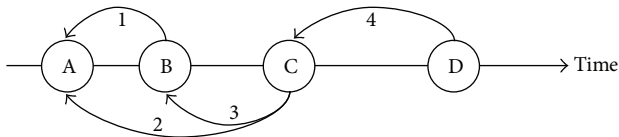


FIGURE 3: Time interval among comments.

4.1. *The Factors of Time.* When reading a review or a comment, the longer the interval to reply is, the weaker its influence is. Hence, the impact of time should be considered [18]. As the above analysis, this section will propose a model to explain the time factor and the comments of the relationship between the strength of influence; this impact of time is shown in Figure 3.

As shown in Figure 3, there is a comment set. The first comments will influence the late comment. For example, “B → A” represents that the comment B is affected by the comment A. The distance between those two comments denotes the time intervals. The larger the interval is, the weaker the influence is [19]. For example, the distance between A and C is greater than the distance between C and B; thus, the impact of the B on the C is greater than its impact

on A. On the other hand, the influence of comment set C will change over time. Therefore, the link weight between comments not only is related to the similarity wt_{ij} but also gradually changes with time [20].

According to the above analysis, it is found that there is an important relationship between the release time of the comment and the choice of the comment. Defining a function f to reduce the probability of the selection of comment, a function of distance on the time is defined as

$$f(t_1, t_2, D) = D^{|t_2 - t_1| \times K}, \quad (3)$$

where f is a function of time t_1, t_2 , and the damping D . The term t_1 is the time of the respondents comment, t_2 is the time when the replying person proposed a comment. Hence, f is time-varying function. If the reply comment is far away from now, the comment has a smaller probability to be accessed. In (3), D ($0 < D < 1$) is a time-dependent coefficient and K is a control factor. Thus, it can choose an appropriate value of K to enlarge or reduce the time. The larger the distance $|t_2 - t_1|$ is, the smaller the impact becomes. Additionally, f changes with time are shown in Figure 4.

As shown in Figure 4, the function f is gradually changing with the time interval, where $D = 0.85$. Thus, the function f is defined reasonably.

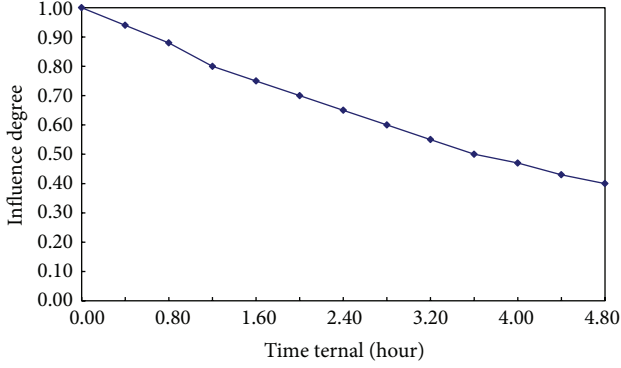


FIGURE 4: The response of the probability function f changing with time.

4.2. *The Detection of the Most Influential Comments.* In a single theme network, if C_i and C_j are explicitly linked, then it leads to $wt_{i,j} = 1$; the impact thus practically exists. If the link relationship is implicit, then $0 \leq |wt_{i,j}| \leq 1$; that is, the impact may exist between comments. For each $wt_{i,j}$, due to the effect of the time can be reduced by using the function f , and then it follows that

$$wt_{i,j}' = f_{i,j} \times wt_{i,j}. \quad (4)$$

To normalize the probability, it needs to normalize the value Sum_i :

$$Sum_i = \sum_{j=1}^n |wt_{i,j}'|. \quad (5)$$

By using (5), the transformed probability can be obtained as

$$wt_{i,j}'' = \frac{wt_{i,j}'}{Sum_i}. \quad (6)$$

Then, the matrix of the improved finite Markov chain can be described as follows.

(a) Step 1

$$A = \begin{bmatrix} wt_{1,1} & \cdots & wt_{1,n} \\ \vdots & & \vdots \\ wt_{n,1} & \cdots & wt_{n,n} \end{bmatrix}. \quad (7)$$

(b) Step 2

$$A' = \begin{bmatrix} f_{1,1}wt_{1,1} & \cdots & f_{1,n}wt_{1,n} \\ \vdots & & \vdots \\ f_{n,1}wt_{n,1} & \cdots & f_{n,n}wt_{n,n} \end{bmatrix}. \quad (8)$$

(c) Step 3

$$A'' = \begin{bmatrix} \frac{f_{1,1}wt_{1,1}}{Sum_1} & \cdots & \frac{f_{1,n}wt_{1,n}}{Sum_1} \\ \vdots & \ddots & \vdots \\ \frac{f_{n,1}wt_{n,1}}{Sum_n} & \cdots & \frac{f_{n,n}wt_{n,n}}{Sum_n} \end{bmatrix}. \quad (9)$$

As shown above, A , A' , and A'' are sequentially linked only considering the time factor and the normalized matrix.

4.3. *The Improved Model of the Finite Markov Chain.* In the field of information retrieval, the PageRank algorithm is widely used. Inspired by this algorithm, a random walk model called Dynamic Opinion Rank is proposed in this section. These algorithms not only take the emotional factors into account but also consider the time factor [21].

From the standpoint view of model use, if a comment gets more positive reviews, then it will be more influential. Moreover, if this comment also replies to other comments, according to the characteristics of the model PageRank, it is reasonable that its influence also will be passed each other. Usually, comments may be affected by the following two cases:

- (1) the comments raised by users are affected by the interested opinion with a probability d ;
- (2) comments may also be subject to random probability $1 - d$ effects.

Based on the above analysis, an algorithm similar to PageRank is proposed as follows:

$$P = \left[(1 - d) \frac{E}{n} + dA^T \right] P, \quad (10)$$

where A is an $n \times n$ improved finite Markov chain transition matrix and n represents the set of comments in the C . Transposing A yields

$$A^T = \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{bmatrix}, \quad (11)$$

where any line of of A^T denoted by A^{Ti} ($1 \leq i \leq n$) represents all the cases that C_j links to C_i ($1 \leq i \neq j \leq n$).

For any element $a_{ji} = (f_{j,i} \times wt_{j,i}) / Sum_j$, the ranking score can be calculated by using the following equation:

$$P(i) = (1 - d) + d \sum_{(j,i) \in E} f(t_i, t_j, D) \frac{P(j) \times wt_{ji}}{Sum_j}, \quad (12)$$

where $P(i)$ represents the authority of the value. Following the above methods, it can eventually obtain the score over a period of time. Then, the comment with the maximum

TABLE 2: The changes of opinion leader in the comments.

Number	Content of comment	Ranking score
Number 25	Comment Number 1	6.581
Number 166	Comment Number 2	6.655
Number 25	Comment Number 3	7.187
Number 25	Comment Number 4	7.119

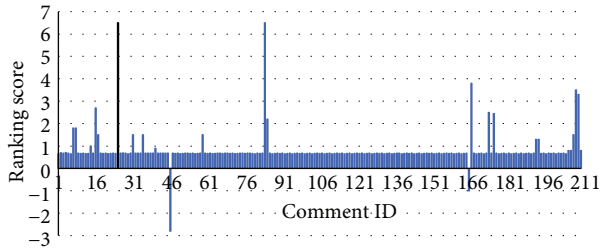


FIGURE 5: Ranking (07:55:17-09:26:46).

authority value can be chosen as the opinion leader. If a comment gets a lot of penetration and most of them are positive emotions consistent link and further the interval time between the comments is not very long, then this comment may get higher ranking scores.

5. Experimental Results and Analysis

To verify the proposed algorithm, experimental analysis is conducted. The data for the experiment is obtained from Weibo news. Through tracking this news within two days (2012-08-17 07:55: 17~2012-08-18 05:42:43) and dividing this period of time into four different time periods, each time period was analyzed to identify opinion leaders, and the dynamic change of the opinion leaders was analyzed.

5.1. The Result of Finding Out the Most Influential Comment.

By building a single-topic comments Network $G_{CN}(V, E)$, setting the parameter $d = 0.85$ and $k = 2$, and applying the algorithm proposed in Section 4, the experimental results are shown in Figures 5 and 6. As shown in Figure 6, there are 211 comments in the first time period. It is also easy to find that the comment Number 25 received the highest scores. However, due to the short time, the relationships between comments are not clear. Therefore, opinion leaders may change with time; it leads to the inaccurate opinion leader. As shown in Figure 6, it is found that Number 25 is not the opinion leader, while Number 166 received the highest score. Hence, Number 166 is opinion leader at this time. Some comments received low scores because their views are not accepted by others. Although there has been an increase of the comment number during the second time of period, the relationship between comments still appears to be relatively sparse.

As illustrated in Figure 7, Number 25 becomes opinion leader, while the score of Number 166 decreases with time. Additionally, due to the increasing number of comments,

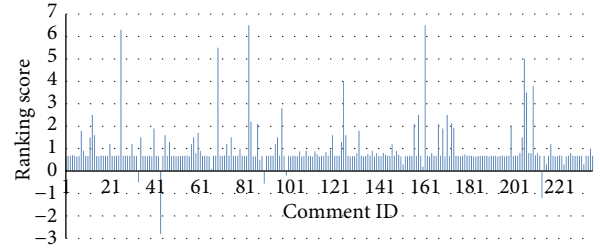


FIGURE 6: Ranking (07:55:17-12:08:44).

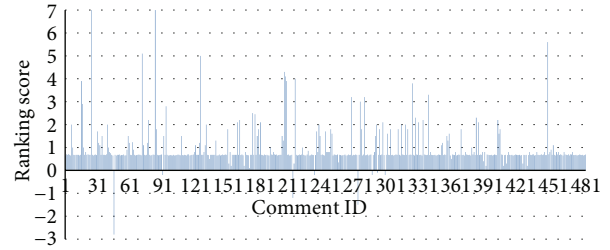


FIGURE 7: Ranking (07:55:17-18:17:19).

the relationship between comments becomes more and more dense and the status of the comments converges to be stable.

It is found in Figure 8 that the number of comments is 560 in the fourth time of period. Now, Number 25 received the highest score, and it is the opinion leader at this time. In comparison with Figure 7, the newly published comment's score grows faster. The result shown in Figure 8 also demonstrates that new comments will get more attention, and it also proves that it is reasonable to take time into account. On the other hand, many comments' scores are growing. Due to the characteristics of news comments, it will get less attention after a period of time. Moreover, the number of comments also will decrease. Hence, the leadership of Number 25 will be maintained for a long time.

The most influential comments and the sort scores are shown in Table 2. We find that opinion leaders are changing over time. Moreover, the rank of opinion leaders is also affected by the time. This also verifies that it is quite necessary to take time into consideration when developing the algorithm.

To evaluate the performance of Dynamic Opinion Rank algorithm, a standard should be proposed to allow the experts to divide comments in each time period into two categories: the strong and weak influence. Then, it needs to measure the time the comment raised, degree centrality, the degree of authority, and the F -Score of several Opinion Rank algorithms. Those comparison results are shown in Figure 9. It is found that the Dynamic Opinion Rank algorithm has much more accuracy and stability than other approaches. It thus verifies the effectiveness of the proposed scheme.

5.2. The Finding of the Most Influential Users. In the process of finding out most influential users, single-topic user network $G_{UN}(V, E)$ should be constructed firstly, and then the

TABLE 3: Clustering results for detection of opinion leader from users.

08:04:37–11:07:11		08:04:07–13:45:49		08:04:37–18:43:22		08:04:37–6:47:08	
Cluster	User count	Cluster	User count	Cluster	User count	Cluster	User count
1	43	1	71	1	117	1	110
2	4	2	11	2	3	2	17
3	2	3	1	3	1	3	1
4	3	4	11			4	
5	1	5					
uid	15928539	uid	15928244		1199491660	uid	15930288

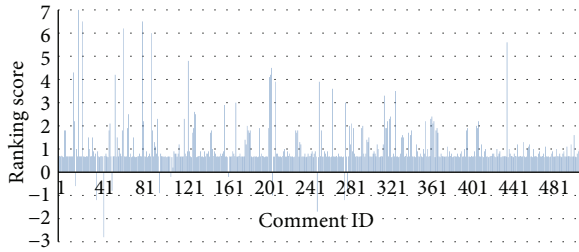


FIGURE 8: Ranking (07:55:17–next day 05:42:43).

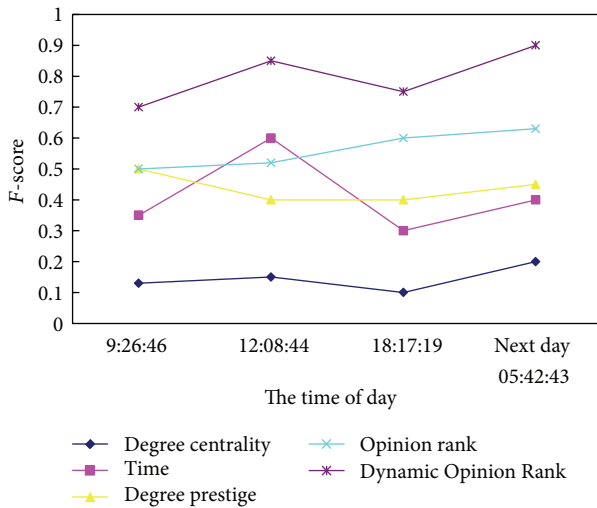


FIGURE 9: F-Score of each method.

proposed algorithm in Section 4 should be applied to detect the most influential users. For the DBSCAN density-based clustering algorithm, the radius range is set between 0.06 and 0.12, and the initial MinPts is chosen as 1. Consequently, there exist some clusters containing noise. With application of the proposed approach, 3~5 clusters are finally obtained, it is shown in Table 3.

From Table 3, we find that the first clusters with most elements can be removed then replaced by clusters with less elements. As $G_{UN}(V, E)$ is sparse, set $w_{PD} = 10$, $w_{CQ} = 0.3$ and $w_{DC} = 10$. Finally, it detects opinion leaders in each period. The result is illustrated in the last line of Table 3. According to the experiment, it reveals that opinion leaders can change with time dynamically.

6. Conclusions

This paper presents a Dynamic Opinion Rank algorithm to find out the opinion leaders in Chinese news. Unlike the existing approaches, the proposed network model explicitly takes explicit and implicit links into account. Moreover, the proposed algorithm was able to conclude that the most influential comments and the opinion leaders were time-varying. Experimental results further verified the effectiveness of the proposed strategy.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

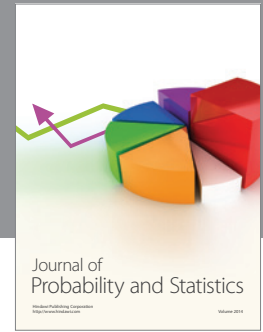
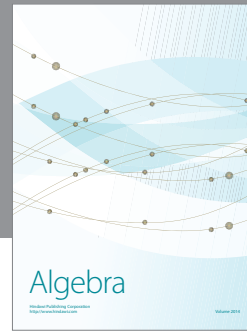
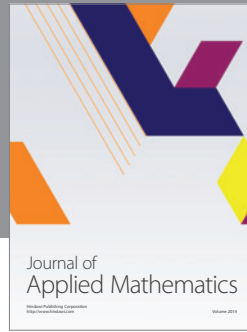
Acknowledgments

This work was also supported by the National Natural Science Foundation of China (Project no. 71171068) and the Polish-Norwegian Research Programme (Project no. Pol-Nor/200957/47/2013). The authors highly appreciate the above financial supports.

References

- [1] M. J. Aschwanden, J. P. Wuelser, N. V. Nitta, J. R. Lemen, M. L. DeRosa, and A. Malanushenko, "first three-dimensional reconstructions of coronal loops with the stereo A+B spacecraft. IV: magnetic modeling with twisted force-free fields," *Astrophysical Journal*, vol. 756, no. 2, article 124, 22 pages, 2012.
- [2] S. Yin, S. X. Ding, A. H. A. Sari, and H. Hao, "Data-driven monitoring for stochastic systems and its application on batch process," *International Journal of Systems Science*, vol. 44, no. 7, pp. 1366–1376, 2013.
- [3] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.
- [4] S. Yin, H. Luo, and S. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2402–2411, 2014.
- [5] S. Yin, G. Wang, and H. R. Karimi, "Data-driven design of robust fault detection system for wind turbines," *Mechatronics*, 2013.

- [6] S. Yin, X. Yang, and H. R. Karimi, "Data-driven adaptive observer for fault diagnosis," *Mathematical Problems in Engineering*, vol. 2012, Article ID 832836, 21 pages, 2012.
- [7] X. Zhao, X. Liu, S. Yin, and H. Li, "Improved results on stability of continuous-time switched positive linear systems," *Automatica*, 2013.
- [8] X. Zhao, P. Shi, and L. Zhang, "Asynchronously switched control of a class of slowly switched linear systems," *Systems & Control Letters*, vol. 61, no. 12, pp. 1151–1156, 2012.
- [9] X. Zhao, L. Zhang, and P. Shi, "Stability of a class of switched positive linear time-delay systems," *International Journal of Robust and Nonlinear Control*, vol. 23, no. 5, pp. 578–589, 2013.
- [10] X. Zhao, L. Zhang, P. Shi, and H. Karimi, "Novel stability criteria for TS fuzzy systems," *IEEE Transactions on Fuzzy Systems*, 2013.
- [11] X. Zhao, L. Zhang, P. Shi, and H. Karimi, "Robust control of continuous-time systems with state-dependent uncertainties and its application to electronic circuits," *IEEE Transactions on Industrial Electronics*, 2013.
- [12] X. Zhao, L. Zhang, P. Shi, and M. Liu, "Stability and stabilization of switched linear systems with mode-dependent average dwell time," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1809–1815, 2012.
- [13] X. Zhao, L. Zhang, P. Shi, and M. Liu, "Stability of switched positive linear systems with average dwell time switching," *Automatica*, vol. 48, no. 6, pp. 1132–1137, 2012.
- [14] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan, "Discovering leaders from community actions," in *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM '08)*, pp. 499–508, October 2008.
- [15] F. Bodendorf and C. Kaiser, "Detecting opinion leaders and trends in online communities," in *Proceedings of the 4th International Conference on Digital Society (ICDS '10)*, pp. 124–129, February 2010.
- [16] Z. Hengmin, D. Zeng, and Z. Changli, "Finding leaders from opinion networks," in *Proceedings of the 2009 IEEE International Conference on Intelligence and Security Informatics (ISI '09)*, pp. 266–268, June 2009.
- [17] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.
- [18] L.-W. Ku and H.-H. Chen, "Mining opinions from the web: beyond relevance retrieval," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 12, pp. 1838–1850, 2007.
- [19] F. Kokkoras, E. Lampridou, K. Ntonas, and I. Vlahavas, "Mopis: a multiple opinion summarizer," in *Artificial Intelligence: Theories, Models and Applications*, pp. 110–122, Springer, 2008.
- [20] L.-W. Ku, H.-W. Ho, and H.-H. Chen, "opinion mining and relationship discovery using CopeOpi opinion analysis system," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 7, pp. 1486–1503, 2009.
- [21] J. Wiebe and E. Riloff, "Creating subjective and objective sentence classifiers from unannotated texts," in *Proceedings of the 6th International Conference (CICLing '05)*, pp. 486–497, February 2005.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

