Weibo Event Detection and Analysis Based on Geographic Coordinates (Postprint)

Authors: Li Jinhua, An Zhongjie

Date: 2017-10-11T00:00:00+00:00

Abstract

Objective: To utilize data mining algorithms to detect valuable event information from massive and complex Weibo data. Method: For representative domestic Weibo platforms, Weibo data with geographic coordinates were efficiently collected through open APIs of the Weibo network. Three data mining algorithms—K-means, KNN, and decision trees—were employed to construct geographic regularity features of Weibo based on five indicators: post count, repost count, comment count, user activity level, and mobility intensity. By comparing the characteristics of daily regional Weibo data with the geographic regularity patterns of Weibo features in that region, events occurring in the area can be detected. Results: Using Weibo data from April 15 and 16, 2015 as test corpus, the proposed Weibo event detection framework successfully detected the "Beijing Sandstorm" event. Limitations: The sample data employed for extracting geographic regularity features of Weibo was relatively limited, which to some extent affected the effectiveness of the event detection framework. Conclusion: The geographic coordinate-based Weibo event detection framework is practical and effective. The analyzed event information can not only help users obtain news about events of interest, but also assist government departments in public opinion monitoring and administrative decision-making.

Full Text

Analyzing Geographical Coordinates Data for Micro-blog Trending Events

Li Jinhua, An Zhongjie
School of Information Management, Central China Normal University, Wuhan 430079, China

Abstract

[**Objective**] This study aims to retrieve trending events from the micro-blog platform with the help of data mining algorithms. [**Methods**] First, we collected micro-blog messages with geographic coordinates from a representative domestic micro-blog website using its open API. Then, we employed three data mining algorithms—K-means, KNN, and decision trees—to construct geographical regularity features based on five indicators: the number of published posts, re-tweets, comments, user activity level, and user movement intensity. By comparing these daily regional micro-blog data features with the established geographical regularity patterns, we detected whether events occurred in specific regions. [**Results**] Using micro-blog data from April 15 and 16, 2015, as test corpora, our proposed event detection framework successfully identified the "Beijing Sandstorm" event. [**Limitations**] The sample data used to extract geographical regularity features was relatively small, which somewhat affected the effectiveness of the detection framework. [**Conclusions**] The geographical coordinate-based micro-blog event detection framework is practical and effective. The analyzed event information can not only help users obtain news about events of interest but also assist government departments in public opinion management and administrative decision-making.

Keywords: Micro-blog; Event detection; Visualization analysis; Geographical coordinates analysis

1. Introduction

In today' s information age, Internet users serve dual roles as both information receivers and creators. The development of the Internet has driven network services toward socialization, transforming them from singular to diversified offerings. Social networks represent various social relationships, organizing strangers, relatives, and colleagues together. Through social networking platforms, users can communicate and interact, forming numerous social circles based on shared values, interests, and beliefs. With the evolution of social networks, micro-blogs have emerged, ushering in a new era of social networking services. Users can disseminate, acquire, and share information through uni-directional and bi-directional follower relationships. Micro-blog users can publish text, images, and videos within 140 characters (including punctuation) via mobile phones, computers, and other clients through mobile WAP services, web browsers, instant messaging software, and SMS, enabling real-time information sharing. Statistics show that 69.0% of micro-blog posts contain images, 8.6% include short links, and some users share videos, enriching and extending micro-blog content. Micro-blog is a real-time broadcasting platform where users can promptly receive information from those they follow. If interested in a particular post, users can comment on or re-tweet it, and they can also broadcast information to their own followers. When a micro-blog is repeatedly re-tweeted, especially by opinion leaders, its dissemination range expands geometrically, creating a "fission-style" information propagation effect.

This paper defines micro-blog events as occurrences posted on the micro-blog platform that trigger extensive user re-tweets and comments, generating significant social impact. While micro-blog research has increased in recent years, studies on micro-blog events remain relatively limited. Domestic and international research on micro-blog event detection primarily focuses on four aspects: sentiment analysis of micro-blog events, dissemination and public opinion control, event detection and tracking, and opinion leader identification. Research on event detection and tracking concentrates on text feature selection, optimization of information similarity calculation algorithms, improvement of topic clustering algorithms, and micro-blog event summarization techniques. The fundamental principle involves word segmentation, feature extraction, and clustering analysis of micro-blog content to mine hot topics. However, unlike traditional web pages and blogs, the brevity of micro-blog content makes it difficult to extract sufficient information to determine whether emergencies have occurred, limiting the effectiveness of content-based emergency detection techniques. This paper uses the mainstream statistical analysis software R to track, acquire, and visually analyze micro-blog hot events from the perspective of geographical distribution data and features.

2. Acquisition of Micro-blog Event Geographical Information

To obtain micro-blog content published by users, we can call Sina Weibo' s `place/nearby_timeline` interface to retrieve dynamic micro-blog information around a specific location. The maximum search radius returned by this interface is 11,132 meters (approximately 11 kilometers), which obviously limits the amount of collectable micro-blog data. To address this issue, we can continuously set different longitude (`long` parameter) and latitude (`lat` parameter) coordinates to collect data near each coordinate point, thereby acquiring sufficient micro-blog content [Figure 1: see original paper].

We used the inscribed square of a search circle to divide the target collection area. Based on the properties of a circle' s inscribed square, if the circumscribed circle radius is 11 kilometers, the inscribed square side length is approximately 16 kilometers. Taking Beijing as an example [Figure 2: see original paper], we illustrate the geographical micro-blog data collection principle. Beijing is located at 115.7°-117.4°E, 39.4°-41.6°N, with an east-west width of approximately 160 kilometers and a north-south length of about 176 kilometers. Therefore, approximately 10 coordinate settings are needed in the latitude direction and about 11 in the longitude direction, requiring roughly 110 coordinate settings to cover the entire Beijing area. Partial data is shown in .

According to the data collection scheme described above, the same micro-blog information may be collected repeatedly, creating data redundancy (as shown in the ABCDEFGHI regions of [Figure 2: see original paper]), occupying substantial unnecessary disk space. Moreover, duplicate data significantly impacts system performance and the effectiveness of event detection. Among micro-blog data attributes, the micro-blog ID is a unique identifier for each post. By

maintaining the uniqueness of this field, we can filter out duplicate micro-blog information to obtain the required experimental data, with partial results shown in .

3. Construction Indicators and Process for Micro-blog Geographical Regularity

3.1 Construction Indicators for Micro-blog Geographical Regularity

To detect micro-blog events, we must evaluate geographical regularity indicators based on collected micro-blog data. This paper focuses on five indicators: micro-blog publication quantity, re-tweet quantity, comment quantity, user activity intensity, and user movement intensity [Figure 3: see original paper]. The rationale for indicator selection is as follows:

(1) **Micro-blog publication quantity**: Driven by public expression motivation, micro-blog users publicly voice opinions and vent emotions about events. Additionally, influenced by social enhancement motivation, many users actively update content to gain more attention, recognition, and followers. Therefore, within specific time periods, the number of micro-blog publications in a region is likely to deviate from normal levels.

(2) **User activity intensity**: This refers to the total number of users who publish at least one micro-blog within a certain time range. After sudden events occur, user activity intensity is likely to deviate from normal levels.

(3) **Micro-blog re-tweet quantity**: Due to various domestic and international factors, numerous social contradictions exist, often leading to man-made emergencies (e.g., terrorism, robbery) and natural emergencies (e.g., earthquakes, floods). Sudden events are the focus of public opinion, and micro-blog users re-tweet posts reflecting such events, causing re-tweet quantities to likely deviate from normal levels. During the Yushu earthquake in Qinghai, one user posted rescue information: "In Yushu' s earthquake disaster area, 100 kilometers west, there is a place called Longbao Town that is severely affected, and currently no rescue teams have arrived." This micro-blog was extensively re-tweeted, eventually attracting government attention and effectively compensating for mainstream media' s information blind spots.

(4) **Micro-blog comment quantity**: For sudden events in a local area, local micro-blog users continuously follow and often express their views, suggestions, and emotions. Additionally, users comment when re-tweeting, resulting in numerous comments on such posts.

(5) **User movement intensity**: Population movement is closely related to sudden events. Movement types include immigration (influx from other regions), emigration (exodus to other regions), and local movement (within the region), with local movement typically reflecting daily routine trajectories. When sudden events occur, user movement intensity falls into an

abnormal range. For instance, during the Yushu earthquake, large numbers of people emigrated to escape the disaster, while during Wuhan University' s cherry blossom season, 100,000-200,000 people flooded into the campus daily, demonstrating that user movement intensity is significant for event detection.

3.2 Construction Process for Micro-blog Geographical Regularity

(1) Temporal distribution of micro-blog features

According to the "2014 Micro-blog User Development Report" released by the Micro-blog Data Center, micro-blog users' publishing, re-tweeting, and commenting behaviors show great similarity over time [Figure 4: see original paper]. The daily trends are as follows: from 0:00-5:00, users are basically sleeping, with all three behaviors declining to low levels. From 5:00-11:00, users gradually become active, with behaviors rising steadily. From 11:00-18:00, behaviors fluctuate slightly but remain at a high, stable level. From 18:00-24:00, users finish work and relax, having sufficient time and energy to publish, re-tweet, and comment, causing behaviors to rise again, peaking at 22:00-23:00. During this period, published posts account for 6.53% of the daily total, re-tweets for 6.37%, and comments for 7.61%. Therefore, we divided daily micro-blog data into these four periods to identify features in each timeframe. User activity intensity and population movement intensity also correlate with daily routines, showing similar temporal distributions.

(2) Spatial distribution of micro-blog features

Due to differences in natural conditions, economic development, and cultural traditions, micro-blog user behaviors vary significantly across regions [Figure 5: see original paper]. Economically powerful and densely populated provinces like Beijing, Shanghai, Guangzhou, and Jiangsu-Zhejiang region have dense and active micro-blog users, generating most micro-blog data. The combined monthly active users of these 10 provinces reach 45.6%.

To reduce the impact of regional differences, we employed the K-means clustering method to minimize intra-class regional differences while maximizing inter-class differences, achieving high cohesion and low coupling. K-means has two defects: the number of clusters K must be predetermined, but the optimal number of categories cannot be known in advance; and initial cluster centers must be manually or randomly determined, with different centers potentially producing completely different results. To mitigate these issues, we propose two solutions: through multiple evaluations of clustering effects on micro-blog data, we derive an experimental K value as the cluster number; and we borrow ideas from K-means++ to address randomness. Considering regional characteristics of micro-blog users, city government locations—being economically developed with high population density and user activity—have the highest probability of becoming cluster centers. Therefore, we can use them as initial cluster centers, then determine the other K-1 centers following K-means++' s principle of maximizing inter-center distances. Using this improved K-means algorithm on data

from and clustering by longitude and latitude, we obtained K cluster centers.

(3) Construction process for micro-blog geographical regularity
After temporal and spatial processing, we reduced the impact of spatiotemporal differences on micro-blog data features. However, some sudden events likely occurred during sample data collection, and eliminating their influence is crucial for summarizing geographical regularity features. We used box plots to exclude outliers. Box plots reflect five statistics: maximum, upper quartile, median, lower quartile, and minimum, while also identifying outliers. Using the standard criterion, we classified data exceeding 1.5 times the interquartile range above the upper quartile or below the lower quartile as outliers. Through box plots of the five indicators, we removed outliers and treated data between maximum and minimum values as normal, calculating normal quantity levels for the five indicators—the geographical regularity features. Compared with [Figure 3: see original paper], [Figure 6: see original paper] shows that micro-blog publication, re-tweeting, commenting, user activity intensity, and user movement intensity all deviate from normal levels, so such outliers must be removed. [Figure 6: see original paper] displays outlier processing results for the 11:00-18:00 period.

For preprocessed micro-blog data, we extracted five-dimensional geographical regularity features: publication quantity by accumulating micro-blog counts in specific spaces across time periods; re-tweet and comment quantities by accumulating all re-tweets and comments in specific spaces across periods; user activity by counting users publishing micro-blogs in specific spaces across periods (activity counted as 1 per user regardless of posts published); and user movement intensity by calculating movement distance from users' consecutive longitude/latitude values and summing these distances.

Using one week of collected data as training samples, we performed K-means clustering to assign all micro-blog data to corresponding classes, obtaining K cluster centers. Data in each cluster were divided into four periods: 0:00-5:00, 5:00-11:00, 11:00-18:00, and 18:00-24:00, yielding geographical regularity features for the five indicators across periods.

4. Micro-blog Event Detection

4.1 Micro-blog Data Boundary Division For daily micro-blog data, we must assign it to corresponding classes for comparison with regularity patterns. Common methods for spatial boundary division include Voronoi Diagrams, which rigidly divide geographical space into polygonal boundaries based on K cluster centers. If micro-blog data coordinates fall within a polygon, the data is assigned to that cluster. However, this algorithm requires comparing numerous boundaries and may assign points very close to one cluster center to another polygon, affecting detection effectiveness. Therefore, we adopted the classic KNN nearest neighbor classification algorithm, using voting to assign data to corresponding classes, achieving logical boundary division.

4.2 Micro-blog Event Detection Process The specific event detection process is as follows:
(1) Use one week of collected micro-blog data with longitude and latitude as input for K-means clustering to obtain K clusters, i.e., K spatial divisions.
(2) Divide data in each cluster into four time periods, removing abnormal time periods from the training set to eliminate interference from already-occurring events.
(3) Extract regularity features for the five indicators (publication, re-tweet, comment, user activity, and movement intensity) from the training sample dataset as comparison standards, storing them in a micro-blog geographical regularity database.
(4) As time progresses, existing regularity features may become outdated, increasing misjudgment probability, so we set an expiration time for periodic reconstruction of regularity features.
(5) Collect daily test data from Sina Weibo through the data collection program and store it locally.
(6) Read data from the local database, use KNN classification to categorize test data, and extract the five indicator features for each cluster.
(7) Compare geographical regularity features from step (3) with daily features from step (6), examining whether indicators deviate from normal levels to determine if major events occurred, issuing alerts when detected.

4.3 Micro-blog Event Classification Based on occurrence processes, mechanisms, and nature, we classify micro-blog events into five categories: natural disasters (e.g., earthquakes, floods), public health incidents (e.g., food safety, animal epidemics), social security events (e.g., terrorist attacks, mass incidents), accident disasters (e.g., environmental pollution, mine collapses), and entertainment/leisure (e.g., celebrity scandals, TV reviews). After identifying clusters with abnormal indicators through the detection framework, we manually read micro-blog content to determine and label event types. When similar events occur in the same cluster, user responses are largely similar. After obtaining a certain amount of labeled cluster data , we can use decision tree classification to predict subsequent event types.

4.4 Micro-blog Event Summarization While the classification framework reveals general event types, summarization is needed to understand specific events. We extract the top 5 hottest micro-blogs as event summaries, proving simple yet effective. Micro-blog heat can be calculated from comment count, re-tweet count, and follower count. More comments indicate active discussion and viewpoint expression. Re-tweeting measures heat dissemination intensity. These metrics correlate with follower count, which follows a power-law distribution where few users have massive followings, making their information more influential. Based on re-tweets, comments, and follower counts, we propose the following heat calculation formula:

$$Hot(W_i) = \alpha R_{wi}^{1/2} + \beta C_{wi}^{1/2} + \kappa \log(FL_{wi} + 1)$$

where α, β, and κ are weight constants (α and β equal 2, κ equals 1), R_{wi} is daily re-tweet count, C_{wi} is daily comment count, and FL_{wi} is follower count. Using this formula, we calculate heat values for each original micro-blog, sort them, and return the top 5 as event summaries.

5. Experimental Results and Analysis

5.1 Data Analysis Results We applied the framework to micro-blog data collected on April 15, 2015. Using KNN classification, we calculated five indicators for micro-blog data and compared them with geographical regularity features, identifying abnormal clusters. During 18:00-24:00, micro-blog publication quantity and user activity exceeded normal levels [Figure 7: see original paper] and [Figure 8: see original paper], while user movement intensity was below normal, and comment and re-tweet quantities were above median levels.

To further understand the event, we calculated heat values using formula (1), sorted them, and returned the top 5 micro-blogs as event summaries . Reading these summaries reveals that Beijing experienced severe sandstorm weather, causing local users to publish numerous original posts tracking the event and expressing dissatisfaction, resulting in above-normal publication quantity and user activity.

[Figure 9: see original paper] shows user movement intensity below normal during 18:00-24:00. As Beijing was experiencing sandstorms with extremely low visibility, primary pollutants shifted from PM2.5 to PM10. Environmental monitoring data shows PM10 concentrations at multiple stations rising sharply after 18:00, with all 35 Beijing monitoring stations exceeding 1,000 g/m^3, reaching severe pollution levels. Under such harsh conditions, Beijing micro-blog users avoided outdoor activities, causing below-normal movement intensity.

[Figure 10: see original paper] and [Figure 11: see original paper] show comment and re-tweet counts above median but within normal ranges. Users who experienced the sandstorm firsthand were more likely to comment on and re-tweet related posts. Driven by information sharing and public expression motivations, users actively commented on and re-tweeted sandstorm posts to express feelings and opinions, causing comment and re-tweet counts to exceed median levels.

5.2 Trend Analysis Improvement To more clearly display the Beijing sandstorm' s development trend, we improved the "topic index" proposed by Chi et al. by adapting it for micro-blog events' shorter cycles. We set hourly intervals and proposed the micro-blog event index (Event Index), defined as a weighting function of hourly micro-blog publication growth relative to the first hour' s growth:

$$EI(E_i) = \frac{(PE_i(t_{j+1}) - PE_i(t_j)) \times P_{base}}{PE_i(t_1)}$$

where E_i represents a micro-blog event, $PE_i(t_j)$ is the cumulative micro-blog quantity related to event E_i from initial time to t_j, $PE_i(t_{j+1}) - PE_i(t_j)$ is the quantity published between t_{j+1} and t_j, $PE_i(t_1)$ is the quantity published in the first hour, and P_{base} is the initial event index (set to 1). Plotting time (hourly intervals) on the x-axis and event index on the y-axis yields a continuous event development trend curve.

Although detected in a single cluster, real-world events affect surrounding areas. Therefore, trend analysis should extend beyond single clusters to the entire affected region. Beijing micro-blog users published 21,966 sandstorm-themed posts on April 15, 2015. By segmenting data hourly and calculating event indices using formula (2) , we visualized the sandstorm trend using R.

[Figure 12: see original paper] reveals that from 0:00-9:00 on April 15, before the sandstorm, only minimal attention existed (incubation period). From 9:00-17:00, the event entered the gestation period with fluctuating indices. From 17:00-18:30, attention rose rapidly as the event' s influence expanded dramatically through the micro-blog platform (acceleration period), coinciding with the sandstorm' s outbreak. From 18:30-19:30, the growth rate slowed, peaking at 2,118.25 at 19:00 (maturity period). Due to micro-blog' s fragmented and superficial attention characteristics, sustained in-depth focus on single events is difficult. From 19:30 on April 15 to 24:00 on April 16, as the sandstorm passed and its impact diminished, user attention shifted to other events (decline period). From 6:00-15:00 on April 16, slight fluctuations occurred as Beijing' s clear skies returned and users contrasted the two days' weather.

6. Conclusion

This paper designs a multi-user micro-blog data collection scheme based on IP rotation for efficient data acquisition. We established five indicators—publication quantity, re-tweet quantity, comment quantity, user activity intensity, and user movement intensity—to measure micro-blog geographical regularity features. Using K-means, KNN, and decision tree algorithms, we proposed a detailed geographical regularity extraction architecture and implemented micro-blog event detection functionality. Experiments verified the method' s effectiveness.

The proposed framework has limitations. Due to experimental constraints, the relatively small sample data used to extract geographical regularity features somewhat affected detection effectiveness, and the visualization module requires manual intervention without full automation. Detecting emergencies nationwide would involve massive data volumes, affecting efficiency. To address these issues, we plan to adopt the Hadoop distributed platform for system construction, deploying HBase on HDFS for big data storage and using MapReduce

for algorithm implementation, enabling efficient nationwide emergency detection. Mahout in the Hadoop ecosystem has already implemented MapReduce versions of our three algorithms, facilitating convenient usage. For visualization, we can use RHadoop (R language combined with Hadoop) to achieve automated processing.

References

[1] Hu Jiming. Study on Open Operation Architecture and Service Expansion of Social Network Service [J]. Information Science, 2012, 30(9): 1396-1400.
[2] Li Biao. The "Portrait Sketch" of Microblogging Opinion Leaders Group —Take 40 Opinion Leaders from Microblogs as an Example [J]. Journalism Review, 2012(09): 19-25.
[3] Yang Liang, Lin Yuan, Lin Hongfei. Micro-Blog Hot Events Detection Based on Emotion Distribution [J]. Journal of Chinese Information Processing, 2012, 26(1): 84-90.
[4] Wang Lin, Shi Kan, Zhao Yang, et al. Experimental Studies on Public Opinion Perception of the Micro Blog's Collective Behavior Based on the Emergencies [J]. Journal of Intelligence, 2013, 32(5): 32-37.
[5] Yang Juanjuan, Yang Lanrong, Zeng Runxi, et al. Research on Communication Mechanism of Internet Public Opinion of Government Affairs Microblog in Public Security Events: A Case Study of the "Shanghai Fabu" [J]. Journal of Intelligence, 2013, 32(9): 11-15.
[6] Lan Yuexin. Research on Microblog Opinion Diffusion Model of Emergent Events [J]. Information Science, 2013, 31(3): 31-34.
[7] Wang Yong, Xiao Shibin, Guo Yixiu, et al. Research on Chinese Micro-blog Bursty Topics Detection [J]. New Technology of Library and Information Service, 2013(2): 57-62.
[8] Chen Guolan. Micro-blog Emergencies Detection Approach Based on Burst Words Distinguishing [J]. Journal of Intelligence, 2014, 33(9): 123-128.
[9] Liu Zhiming, Liu Lu. Recognition and Analysis of Opinion Leaders in Microblog Public Opinions [J]. Systems Engineering, 2011, 29(6): 8-16.
[10] Wei Zhihui, He Yue. Identify Microblogging Opinion Leaders Based on Information Entropy and Unascertained Measure Model—Taking "Emergencies of Qingyang School Bus" as an Example [J]. Information Science, 2014, 32(10): 38-43.
[11] Tian Ye. On Trends Analysis and Prediction Based on Micro-Blogging Platforms [D]. Wuhan: Wuhan University, 2012.
[12] Yang Y, Carbonell J, Brown R. Multi-Strategy Learning for Topic Detection and Tracking [A]. // Topic Detection and Tracking [M]. Springer, 2002: 85-114.
[13] Feng Yong, Han Nan, Jia Dongfeng. Microblog Events Detection and Tracking with Incremental Hierarchical DBSCAN Based on Representative Posts Using Cloud Framework [J]. Journal of Computer Applications, 2013, 33(12): 3559-3562.
[14] Wang Lianxi. A Literature Review on Pre-processing and Learning of Mi-

crotext [J]. Library and Information Service, 2013, 57(11): 125-131.
[15] Fu C, Samet H, Sankaranarayanan J. WeiboStand: Capturing Chinese Breaking News Using Weibo "Tweets" [C]. In: Proceedings of the 7th ACM SIGSPATIAL Workshop on Location-Based Social Networks. 2014.
[16] Wang Feng. "Micro" Forces of the Catastrophic Event—Qinghai Yushu Weibo Application Analysis in the Earthquake [J]. News World, 2010(S2): 149-150.
[17] Zhang P. Social Inclusion or Exclusion? When Weibo (Microblogging) Meets the "New Generation" of Rural Migrant Workers [J]. Library Trends, 2013, 62(1): 63-80.
[18] Weibo Data Center. The 2014 Report of Weibo Users Development [R/OL]. [2015-02-06]. http://data.weibo.com/report/reportDetail?id=215.
[19] Wu Suhui, Cheng Ying, Zheng Yanning, et al. Survey on K-means Algorithm [J]. New Technology of Library and Information Service, 2011(5): 28-35.
[20] Qi Feng, Liu Kun, Zhang Chao, et al. A Novel Base Station Coverage Simulation Based on Intersection of Circle and Voronoi [J]. Journal of Beijing University of Posts and Telecommunications, 2014, 37(S1): 108-114.
[21] Jiang Tao, Chen Xiaoli, Zhang Yufang, et al. Improved KNN Using Clustering Algorithm [J]. Computer Engineering and Applications, 2009, 45(7): 153-158.
[22] Lu Ansheng, Chen Yongqiang, Tu Haowen. The Analysis and Application of Decision Tree Algorithm of C5 [J]. Computer Knowledge and Technology, 2005(3): 17-20.
[23] Chi Chengying, Li Hong. Network News Hot Topics Detection and Tracking Based on Modified TF*PDF Algorithm [J]. Computer Applications and Software, 2013, 30(12): 311-314.
[24] Xie Kefan, Zhao Shi, Chen Gang, et al. Research on Lifecycle Principle and Group Decision-making of Network Public Sentiment Emergency [J]. Journal of Wuhan University of Technology: Social Sciences Edition, 2010, 23(4): 482-486.
[25] Narayanam R, Narahari Y. A Shapley Value-based Approach to Discover Influential Nodes in Social Networks [J]. IEEE Transactions on Automation Science and Engineering, 2011, 8(1): 130-147.
[26] Chen Jirong, Le Jiajin. Reviewing the Big Data Solution Based on Hadoop Ecosystem [J]. Computer Engineering & Science, 2013, 35(10): 25-35.

Author Contributions

Li Jinhua: Proposed and demonstrated the research proposition, partially wrote and revised the final version.
An Zhongjie: Collected and analyzed data, proposed experimental verification plans, and revised the paper.

Conflict of Interest Statement

All authors declare no conflict of interest.

Supporting Data

Supporting data [1-2] are stored by the authors and can be obtained via email: anzhongjie01@163.com. Supporting data [3-11] are available in the online version of the journal at http://www.infotech.ac.cn.

Note: Figure translations are in progress. See original paper for figures.