Research on the mining of opinion community for social media based on sentiment analysis and regional distribution

Baocheng Huang1, Guang Yu2

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

E-mail: huanghljhrbin@gmail.com

2. School of Management, Harbin Institute of Technology, Harbin, China

E-mail: yug@hit.edu.cn

Abstract: Many researches have been carried out on network comments nowadays. In order to get more sentiment types, the paper put forward a opinion community discovery method based on sentiment of web comments and regional distribution, considering that Internet users from different regions have different perspectives on the same issue. First, the research clusters network comments into different opinion communities according to sentiment similarity. Then in the paper the Longest Sequential Sentiment Phrase (LSSP) is defined. The representative view of each pinion community is extracted by the mining algorithm of LSSP. At last, it studies regional distribution of opinion communities. Experimental results show that the method can be applied to analyze network comments. The research also shows that we can get more information from network comments than other related network public opinion analysis.

Key Words: opinion community; sentiment clustering; regional distribution

1. INTRODUCTION

Public opinion is comprehensive views of personal attitudes and beliefs held for adults. Opinion involves sentiment. According to different sentiment propensity of Internet users, any opinion can be classified to positive, negative and neutral. However, public opinion is more complex than sentiment propensity. Public opinion involves sentiment. However, the sentiment cannot be divided into "positive, negative and neutral" simply. More sentiment should be considered, such as hover, Depressed , expectancy, Hopeful and so on. [1] The opinion expressed may be "condemned", "helpless", "praised", "blessing" and so on. Traditional sentiment propensity analysis can only be two categories or more categories. It cannot get deep-seated sentiment characteristics. Traditional sentiment propensity analysis can only be. It cannot get deep-seated sentiment characteristics. To get accurate sentiment from the public opinions, it needs to analyze the latent semantic of comments. Abstract the latent sentiment semantic involved in the comments. [17-24]

Web2.0 provides different space and tools, web comment forms and users' expression. Different people on the same event may hold different views. That led to a variety of network media content. It is an urgent problem to automatically collect and analyze the user's attitude on various issues. On the Internet, users can express their views and opinions on a particular event or merchandise.[25-31] People who published similar sentiments can form a virtual community. It can be called the opinion community. By analyzing the comments' sentiments, aggregate comments with similar sentiments into several different groups. In this method, we can get different opinion communities representing different sentiment propensity. [2]

This work was supported by the National Natural Science Foundation of China (Project no. 71171068). The authors highly appreciate the above financial supports.

2. MINING MODEL OF OPINION COMMUNITY

Through the analysis of previous section, opinion community is virtual community constituted by network Comments which Contain Similar motion. It is an effective method to get more fine-grained analysis of network comments. Building process of opinion community involves the following steps:

2.1 Access to network Comments data

For an event, each portal will report related events and the user will publish relevant Comments. However, due to different sites with different Build Framework, we need to analyze comments' URL according to different websites first. Then abstract Comments By reptiles tool, analyze and stored it into text format for subsequent analysis. The data is from Sina API. The maximum allowed number of pages is 6.

2.2 Comments text modeling

For each comment text, establish vector space model through feature extraction. However, it is need to analyze the comments Emotion of the content which is potential information contained in the content. So we shall analyze the Latent Semantic contained in the comments and extract the underlying emotional semantic features.

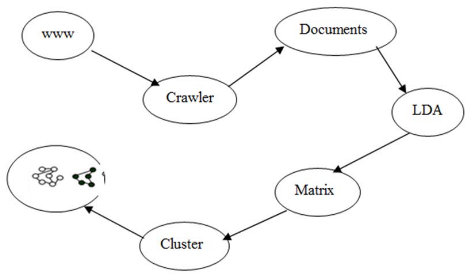

Fig. 1 Work flow chart of discovering opinion community

2.3 find the opinion community

By modeling the comment text, it is expressed by Vector. Then group texts with similar emotion content by Clustering technology. In this method, we can several text Clusters. The texts of the same Clusters have similar motion. The cluster is a representation of opinion community. The whole process of opinion community found can be expressed by the following figure.

3. REPRESENTATIVE PERSPECTIVE OF OPINION COMMUNITY EXTRACTION MODEL

3.1 Representative perspective of opinion Community extraction

opinion comments can be divided into different clusters by the opinion Community finding model in the last Section.

As each cluster is grouped according to the similarity of the emotion. Each cluster can be called an opinion Community. As each cluster Contains similar view. Each cluster can abstract a Representative view.

For each cluster, such as opinion Community cl_i (i=1, 2, …, n), we define op_1, op_2, … as Representative view of opinion Community cl_i.[3]

A representative view may be an emotional phrase. Therefore, representative view of cli can be marked as cl_i. opk (k = 1, 2, ...).

As each cluster is grouped by similar motion, each representative view of a cluster is Independent. Therefore, the representative view of a cluster shall appear frequently in the cluster, but not in the other clusters. It is considered that a cluster may have several representation views which appear frequently. In order to express the representation view accurately, we choose the Longest one as the final representation view of the cluster's emotion. [4]

3.2 Longest Sequential Sentiment Phrase

Definition 1: For each opinion community cl_i, pr represent emotional Sentences, phrases or words of a comment cm.ct (cm $\in$cl) in the opinion Community.[5]

We define the amount of words in pr as pr's Length. If the length of pr is i, pr is marked by i-phrase.

For any i-phrase (i>1), e.g. i=2, the phrase composed of word1word2 is different from the phrase word2word1. The phrase is called a sequential emotional phrase and marked as SSP.

SSP-Set represents a SSP group. And all the emotion phrases are involved in sequential emotional phrase pr$\in$ SSP-Set.

If pr meet the condition that there is none pr'$\supset$pr and pr'$\in$ SSP-Set, we call pr Longest Sequential Sentiment Phrase.[6] Mark it as LSSP.

One Longest Sequential Sentiment Phrase (LSSP) involves the following two characteristics.

(1) LSSP is sequential;

(2)LSSP is the longest emotion phrase;

The paper abstract LSSP from opinion community as the Representative views.

As sequence of words is different, the meaning which Sentences express is different. For example, "Obama supports" and "support Obama" have different meaning. So we emphasize sequence.

What's more, the length of Sentences is different. Long Sentence can express the meaning more exactly. For example, "support Obama win the Nobel Peace Prize" is more exact than "support Obama". So we emphasize the length of Sentence.

3.3 The implementation of opinion extraction based on Sentiment Phrase Tree

In order to abstract Longest Sequential Sentiment Phrase (LSSPs) from Sentiment Phrase group, the paper use preamble tree to represent Sentiment Phrase group. We call it Sentiment Phrase-tree (SP-tree).

Definition 2 (SP-tree) ： sentiment Phrase-tree contains one root node, one Sentiment Phrase preamble tree set , as the child node of the root node and an index table of words.[7]

On the Sentiment Phrase-tree, each node express one word of Sentiment Phrase except root node.

From the root node to the leaf node, the words on one route can express a Longest Sequential Sentiment Phrase.

Each node is composed of word, word frequency , node pointer. Word frequency represent the amount of Sentiment phrases. These Sentiment Phrases have such a feature that the path of the portion through the node. Node pointer points to the node of next word of Sentiment phrases. SP-tree contains the words that are included in the node. If there is no next node, the pointer is null.[8]

Each word in the index table is the first node of a preamble sub-tree .

We can abstract Frequent Longest Sequential Sentiment sentence from any Sequential community according to Sentiment Phrase-tree.

For each word in the indexes table of Sentiment Phrase-tree, we can get Sentiment Phrases with different length and calculate their frequency according to node pointer and sub-tree.

Base on frequency threshold and Phrases' length, we can find all Longest Sequential Sentiment Phrases from Sentiment Phrase-trees. Algorithm 1 describes the finding Process.

```
Algorithm 1: find LSSPs from SP-tree
Input: SP-tree of a certain sentiment community. frequency threshold is θ
Output : LSSPs of the input sentiment community
Method:
1)SSP-Set={};
2)for all word wd in word index table of SP-tree;
3){find node-link of wd;
4)find subtree of wd and compute count of wd+subtree recursively;
5)if count of wd+subtree satisfies θ then SSP-Set=+{wd+subtree};
6)}; find the LSSPs from SSP-Set;
```

As it is need to find Longest Sequential Sentiment Phrases on Sentiment Phrase-trees only, the Algorithm can be optimized compared with frequent pattern mining algorithm and maximal frequent pattern mining algorithm.

This optimization is not specifically considered in the paper. By abstracting process, we can get LSSPs and the frequency appeared in the certain opinion community.

Referring to thought of TFIDF (Salton, Buckley, 1988) [9], we treat the LSSPs which appear frequently in the abstracted opinion community and infrequently in others as representation view of the certain opinion community.

We define a vector LFICF to represent weight of each lssp.

Calculation Method as shown below :

$$W(lssp_i, j) = LF_j(lssp_i) * ICF_j(lssp_i)$$

These are defined as wherein

$$LF_j(lssp_i) = frquency(lssp_i) / \max\{frquency(lssp)\}$$

$$ICF(lssp_i) = \log(N / n(i))$$

For a LSSP of the certain Cluster cl, $LF_j(lssp_i)$ is the frequency that $lssp_i$ appears in the opinion community cl_j. $ICF(lssp_i)$ is the number of the inverse clusters of the total containing $lssp_i$ clusters.[10]

We can get the Calculated LFICF value of each LSSP in a certain cluster. Then sequence the LFICF values. Abstract the first k LSSPs as the representative views of the cluster.

4. IMPLEMENT OF OPINION REGIONAL DISTRIBUTION ANALYSIS ALGORITHM

For the comments we get, it can be found that all addresses of comments are represented by cities though Analysis. In order to show the addresses on the map Broadly, we Classify the addresses to Province before Clustering process.

Algorithm2 shows the process of opinion regional distribution

```
Algorithm2 opinion regional distribution analysis
Input: a opinion community cl, regional concept hierarchy tree R-tree
Output : regional distribution of the opinion community cl
Method;
1)for all cm∈cl;
2){if cm.ad is in R-tree
3)if cm.ad is not in province level
4)province=getFather(cm.ad);
5)climb cm.ad to province;
6)if cm.ad is in province level
7)climb cm.ad to province;
8)else accept cm.ad by dialog box;//man-machine conversation
9)}; cluster cm by cm.ad;
```

During the algorithm process, comments' addresses are classified to province level from step 1) to 7).

Though the addresses classification process, the comments of a opinion community can be clustered. Comments of the same province level addresses are clustered into a group.[11]

In Algorithm 2, province level addresses can be returned though city level addresses analysis by line 4). As most comments' addresses are city level, it is the main step of Algorithm 2. Algorithm 3 shows the Implementation process.

```
Algorithm 3: return the province addresses according cities.
Input : a comment's address of city level cm.ad
Output: the province level address of the comments
Method:
1) i=1;
2)forPi in N;
4){ j=1;
5)for cj in Pi -> child list;
6){ if(cj==cm.ad)
7)return Pi
8)else
9)j++;
10)};
11)i++;
10)};
11)else accept cm.ad by dialog box; //man-machine conversation
```

By Algorithm 3, we can get province level address according to comments' city level address.

In this method, all comments' addresses can be Classified into province level. Then cluster the comments of a opinion community. We can get the regional distribution of a certain opinion community. It can be helpful to analyze the comments meticulously.

5. EXPERIMENT RESULTS AND ANALYSIS

In order to verify the research in this paper, we select network comments of hot events in the last three years as the experiment data. The experiment data in the paper are from Sina Weibo. We analyze the comments about news "Obama win the Nobel Peace Prize", "Spring Festival Evening is Vulgar" in 2012 and "Bye! Bin Laden!" in 2011.

5.1 Results and analysis about opinion community discovery based on LDA+Kmeans

In the process of LDA+kmeans opinion community discovery analysis, it is first to classify and Remove stop words of network comments which are downloaded from internet.[12] Write all valid comments into text by lines in accordance with the input data form which LDA Latent Semantic Analysis need.[13]

According to LDA Latent Semantic Analysis model, we set 10 topics among the comments of our Experiment. Each comment text is dealt with 2000 times iteration. Abstract the first 10 words which are highest weight after iteration

We can get matrices θ , γ and 10 highest weight topics words after LDA iteration. They are stored as text.

Each text relate to a vector θ. Each item in the vector θ represents the probability that the semantics of text generation.

The values of each row in the matrix represents the involved degree of each semantic which contained in document d.

The values of each column in the matrix represents the involved degree of a certain semantic which contained in all documents.

M θs relate to M texts. If we gather the M vectors to build up a matrix, it represents the same meaning with matrix γ.

So matrixγ is selected to do k-means cluster analysis. Take the clustering results as opinion community related to sentiment information which network comments contain.[14]

Though the Pretreatment of comments text, we filter out the invalid part of comments. The comments texts used for analyzing can represent some sentiment information.

There are 4250 valid comments related to the event “Obama win the Nobel Peace Prize”. By LDA Latent Semantic iteration Analysis, we can get Semantic information as table 1.

In the table, each row in the matrix represents a sentiment topic. The table shows the first 10 highest weight topic words after iteration. These words are recapitulative Description of topic sentiment.

TABLE 1 Top 10 words of each topic

Topic	Top 10 words
1	World peace too dizzy comment make great maintenance black
2	People support world point feel quot sad this qualification in
3	America china human develop country in human problem hope
4	laugh really do too day joke floor death before irony
5	China college America words one world backwards life say conscience
6	Peace Prize Obama war get Afghanistan Iraq aggression kill
7	Nobel country most now leader do already western really a
8	United States say know good only want invention judge country no
9	Chairman hu award shahaha China money support next eat again
10	Prize should Obama award Laden little send Bush little was promise

Cluster the Semantic vector γ in the method of k-means after iteration. Clustering results as Figure 2 shows. Each histogram in the figure is statistics of comments contained in the opinion community. It does not represent Specific content.

We can get the effect of each clustering from the figure. The clustering effect can reflect the number of sentiment categories based on sentiment clustering.

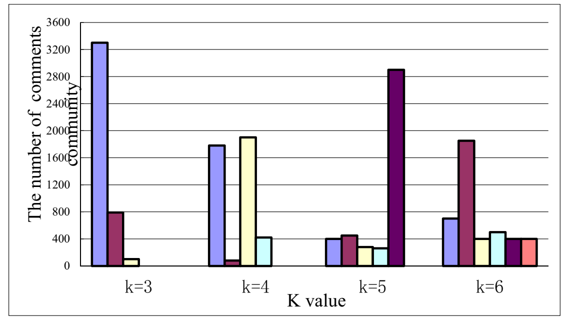

Fig.2 The different results of k-means about “Obama Win Nobel Prize”

There are 7438 valid comments related to the report “Spring Festival Evening is Vulgar”.

After LDA latent semantic iteration analysis, we can get semantic information as table 2 shows.

TABLE 2 Top 10 words of each topic

Topic	Top 10 words
1	CCTV Spring Festival Gala too Zhao now party become should problem already
2	Two-person show like now people know good things local art audiences
3	Spring Festival Gala program director advertising this year getting really feeling too laugh
4	small pieces Benshan Zhao Shenyang humor really play this year most
5	elegant art level people vulgar appreciate enjoy the party high quality should
6	vulgar good nationwide support people CPPCC members Jin Tielin garbage
7	Culture China peasant folk seed society art representatives nation
8	people money North East North South Southerners eat really northerners really
9	people want to know a China curse only listen to think again curse again
10	say support Jin teacher good look please next sentence

In the process of analyzing sentiment about network comments “Obama Win Nobel Prize”， we get the first 10 words of 10 sentiment Semantic by LDA Latent Semantic Analysis model.

Analyze the first 10 highest weight words with sentiment Semantic as table 1 shows. They describe three sentiment tendencies mainly: “it is a joke that Obama was awarded the Nobel Peace Prize”, [15] “Grieve! It is ironic for the Nobel Peace Prize”, “The Nobel Peace Prize” is the instrument that the Western powers control the west countries.”

Analyzing the result after clustering, it reflects that the clustering result is better with the condition K=3. This is also in line with the previous LDA iteration results of the latent sentiment semantic.

According to the result of latent semantic LDA iteration in table 2, The words describe for sentiment tendencies mainly: “CCTV Spring Festival Gala has problem itself.”, “Spring Festival Evening is increasingly commercialized”, ”Elegant art is not need in the Spring Festival Gala because not all audience can appreciate it”, ”People of south region do not taste the art of North region, especially Northeast art ”.

From Figure 2, we can know that clustering effect is different with different K values. However, each cluster will gather the information to four major clusters obviously.

The sentiment community is Objective and reasonable with the condition K=4 in process of clustering. It meets the sentiment semantic analysis result.

5.2 Result of the regional correlation of opinion distribution and its analysis

Some public opinions relate with the opinion holders’ regional position and identity. Cultural diversity exists among different regions and results in different altitude standards, which turn out in different opinions towards the same event.[16] In the experiment we verify the regional distribution reasonable through the analysis of the opinion holder’s regional position in opinion community of each event, using tree of the China regional concept layer tree.

When classifying the reviewers’ position to province level, we dismiss TW, HK and MC due to the comments we obtain about a certain event are from the mainland only, and all 31 provinces of the mainland are taken into consideration. We carry out the analysis of opinion community distribution, when K=3,” Obama Win Nobel Prize”; K=4, “the vulgar Spring Festival Gala”, “Bye! Bin Laden!”.

Based on the China regional concept layer tree, we conclude that: (Shown in Table.3)

By analyzing the opinion communities’ regional distribution of the event “Obama achieved Nobel Peace Prize”, we find that most of Chinese people believe that “US controls the Nobel Peace Prize” and “He declared the war and achieved the Peace Prize-that makes no sense”, which suggests “too much benefit in the Noble Prize, it’s a tragedy”. People admit the value of the Noble Prize, but feel miserable when the Noble Prize is controlled to be a tool for benefit by the EU and US.

By analyzing the event “the vulgar Spring Festival gala”, we can find people admire with the Spring Festival gala more or less, they hold that if you like it just enjoy it, be tolerant with it. On the other hand, people disagree with the flood of advertisements and the splendid appearance. Furthermore, cultural diversity exists between the south region and the north region. Northerners, especially the northeast think: Spring Festival gala should represent the national art. Although the couple dance opera is bit of vulgarity, it’s also national art. Spring Festival gala should be common, not vulgar.

By observing the opinion community distribution of the event “Bye! Bin Laden!”, We conclude that Chinese people all the mainland around are against the hegemony of America. They believe “Imperialist hegemony is more horrible than terrorists”. Some feel sympathy about Bin ,”It’s a pity, but Bin is the world idol, he’s a good guy worldwide”. At the same time, people recognize Bin Laden fight against aggression to maintain world peace with radical measures.

Obviously, culture and economy of a certain region do affect the locals’ mind and action. Because of the different local language, it’s difficult to accept the northern art. But this kind of art does not lose popularity. In the open and well-developed Eastern China, people are more open-minded with easy-contact and willing to express the opinions upon hot issues.

6. CONLUSION

This paper demonstrates a opinion discovery method based on sentiment of Web reviews and regional distribution. First, analyze the sentiment of comments and clustering the comments into different opinion communities according to different sentiment. Then abstract the Longest Sequential Sentiment Phrases as the representative view of opinion communities due to the frequency, sequence and length of the sentiment phrases. At last, analyze regional distribution of each sentiment community. The experiment proves that we can abstract more sentiment and information from the network comments.

REFERENCES

[1] Tony V. Word Net sits the SAT: A knowledge-based approach to lexical analogy [A].In Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004)[C], 2004:606-612

[2] Ku, L, Chen, H. Mining Opinions from the Web: Beyond Relevance Retrieval [J], Journal of the American Society for Information Science and Technology, 2007, 58(12): 1838-1850

[3] Tang, H, Tan, S, Cheng, X. A survey on sentiment detection of reviews [J], Expert Systems with Applications, 2009, 36: 10760-10773

[4] Melville, P, Gryc, W, Lawrence, R. Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification[A], KDD[C] 2009:1275-1283

[5] Tan, S, Zhang, J. An empirical study of sentiment analysis for Chinese documents [J], Expert Systems with Applications, 2008, 34: 2622-2629

[6] Tan, S, Wang, Y, Cheng, X. Combining Learn-based and Lexicon-based Techniques for Sentiment Detection without Using Labeled Examples[A], SIGIR [C] 2008: 743-74

[7] Hu, Y, Li, W, Lu, Q. Developing Evaluation Model of Topical Term for Document-Level Sentiment Classification[A],PRICAI[C] 2008, 5351:175-186

[8] Ye, Q, Zhang, Z, Law, R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches[J], Expert Systems with Applications,2009, 36: 6527-6535

[9] Pang, Bo and Lillian Lee, and Vaith yana than, S. Thumbs up? Sentiment classification using machine learning techniques [A], In Proceeding of EMNLP 2002[C], 2002, 79-86

[10] Turney, Peter D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews

[A], In Proceedings of the 40thAnnual Meeting of the Association for Computational Linguistics (ACL) [C], 2002: 417-424

[11] Feng, S, Wang, D, Yu, G, Yang, C, Yang, N. Chinese Blog Clustering by Hidden Sentiment Factors[A], ADMA[C] 2009, 5678: 140-151

[12] Santos, R, He, B, Macdonald, C, Ounis, I. Integrating Proximity to Subjective Sentences for Blog Opinion Retrieval[A], ECIR[C] 2009: 325-336

[13] Gerani, S, Carman, M, Crestani, F. Investigating Learning Approaches for Blog, Post Opinion Retrieval [A]. ECIR[C] 2009: 313-324

[14] Seki, Y, Kando, N, Aono, M. Multilingual opinion holder identification using author and authority viewpoints[J], Information Processing and Management, 2009, 45: 189-199

[15] Missen, M, Boughanem, M. Using WordNet's Semantic Relations for Opinion Detection in Blogs[A], ECIR[C] 2009: 729-733

[16] Kokkoras, F, Lampridou, E, Ntonas, K, Vlahavas, I. MOpiS: A Multiple OpinionSummarizer[A] SETN[C] 2008: 110-122

[17] X. Li and H. Gao. A new model transformation of discrete-time systems with time-varying delay and its application to stability analysis. IEEE Trans. Autom. Control, 56(9):2172--2178, 2011.

[18] X Xie, S Yin, H Gao, O Kaynak. Asymptotic stability and stabilisation of uncertain delta operator systems with time-varying delays. IET Control Theory & Applications, 7(8):1071-1078,2013.

[19] Xudong Zhao, Xingwen Liu. Improved Results on Stability of Continuous-Time Switched Positive Linear Systems. Automatica. 50(2): 614–621, 2014.

[20] H. Gao, T. Chen, J. Lam. A new delay system approach to network based control. Automatica. 2008, 44: 39-52.

[21] F. Liu, H. Gao, J. Qiu, S. Yin, T. Chai, J. Fan, Networked multirate output feedback control for setpoints compensation and its application to rougher flotation process, IEEE Transactions on Industrial Electronics, 61(1):460-468, 2013.

[22] S. Ding, P. Zhang, S. Yin, E. Ding , An integrated design framework of fault-tolerant wireless networked control systems for industrial automatic control applications, IEEE Transactions on Industrial Informatics, 9(1): 462-471, 2013.

[23] X. Li and H. Gao. A Heuristic Approach to Static Output-Feedback Controller Synthesis with Restricted Frequency-Domain Specifications. IEEE Trans. Autom. Control, 2014. DOI:10.1109/TAC.2013.2281472.

[24] Xudong Zhao, Lixian Zhang, Peng Shi, Hamid Reza Karimi. Novel Stability Criteria for T-S Fuzzy Systems. IEEE Transactions on Fuzzy Systems. 21(6):1-11, 2013.

[25] S. Yin, H. Luo, S. Ding, Real-time implementation of fault-tolerant control systems with performance optimization, IEEE Transactions on Industrial Electronics, 64(5):2402-2411, 2014.

[26] S. Yin, G. Wang, H. Karimi, Data-driven design of robust fault detection system for wind turbines, Mechatronics, DOI:10.1016/j.mechatronics. 2013.11.009, 2013.

[27] S. Yin, S. Ding, A. Haghani, H. Hao. Data-driven monitoring for stochastic systems and its application on batch process, International Journal of Systems Science, 44(7):1366 -1376, 2013.

[28] S. Yin, S. Ding, A. Haghani, H. Hao, P. Zhang. A comparison study of basic datadriven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. Journal of Process Control, 22(9):1567-1581, 2012。

[29] S. Yin, X. Yang, H. Karimi, Data-driven adaptive observer for fault diagnosis, Mathematical Problems in Engineering, Volume 2012, Article ID 832836, 21 pages, doi:10.1155/2012/832836, 2012.

[30] S Ding, S Yin, K Peng, H Hao, B Shen, A novel scheme for key performance indicator prediction and diagnosis with application to an industrial hot strip mill, IEEE Transaction on Industrial Informatics 9 (4):2239 – 2247, 2013.

[31] Study on modifications of PLS approach for process monitoring
S Yin, SX Ding, P Zhang, A Hagahni, A NaikIFAC World Congress, pp.:12389-12394, Milano, Italy, 2011.

TABLE 3 REGION DISTRIBUTION OF EVERY OPINION COMMUNITY

	Obama Nobel Prize			Bye! Bin Laden!				the vulgar Spring Festival gala			
	cl1	cl2	cl3	cl1	cl2	cl3	cl4	cl1	cl2	cl3	cl4
Beijing	103	19	3	10	56	11	8	159	71	138	38
Shanghai	88	23	5	1	40	10	3	59	44	54	15
Tianjin	26	4	0	4	18	6	4	35	16	45	11
Chongqing	61	12	3	4	11	2	4	48	25	36	8
Hebei	98	22	2	7	54	13	8	81	36	74	15
Shanxi	54	12	0	7	41	7	9	60	30	42	10
Neimeng	10	4	1	3	19	6	3	35	14	25	5
Liaoning	78	13	1	11	50	14	7	211	67	167	43
Jilin	39	8	4	3	21	6	5	122	28	71	35
Heilongjiang	46	16	2	5	30	17	5	121	46	88	24
Jiangsu	183	37	9	17	89	18	12	129	70	103	35
Zhejiang	229	61	13	22	108	36	24	213	138	140	42
Anhui	89	26	6	8	40	13	5	59	29	56	11
Fujian	166	29	10	14	92	25	9	95	54	66	16
Jiangxi	93	27	5	2	25	9	4	38	21	33	11
Shandong	126	26	9	16	66	24	6	104	52	126	23
Henan	166	40	14	20	68	21	12	95	60	100	26
Hubei	150	35	9	5	47	8	4	80	48	84	11
Hunan	121	33	7	15	74	18	8	53	41	60	17
Guangdong	486	109	24	80	285	71	49	218	149	153	41