Summer 6-30-2018

A Net Loan Monitoring Platform for University Students Based on Visual Micro-blog

Ruijun Zhang
Center of Service Science and Engineering, Wuhan University of Science and Technology， Wuhan, 430065, China;School of Management, Wuhan University of Science and Technology, Wuhan, 430065, China

Caiyan Lin
School of Management, Wuhan University of Science and Technology, Wuhan, 430065, China

Xinxin Zheng
School of Management, Wuhan University of Science and Technology, Wuhan, 430065, China

Lixin Gao
School of Management, Wuhan University of Science and Technology, Wuhan, 430065, China

Jiaxin Tong
School of Management, Wuhan University of Science and Technology, Wuhan, 430065, China

See next page for additional authors

Authors
Ruijun Zhang, Caiyan Lin, Xinxin Zheng, Lixin Gao, Jiaxin Tong, and Shangyang Yu

A Net Loan Monitoring Platform for University Students Based on Visual Micro-blog

Ruijun Zhang[1,2], Caiyan Lin[2], Xinxin Zheng[2], Lixin Gao[2], Jiaxin Tong[2], Shangyang Yu[3]

[1]Center of Service Science and Engineering, Wuhan University of Science and Technology，Wuhan, 430065, China

[2]School of Management, Wuhan University of Science and Technology, Wuhan, 430065, China

[3]College of Science, Wuhan University of Science and Technology, Wuhan, 430065, China

Abstract: In order to prevent college students from tragic events, which mostly result from falling into illegal net loan, a net loan risk monitoring platform of college students was established based on the microblog visualization. Based on the calculation model of risk degree (CMRD) and the calculation model of relational closeness (CMRC), a visual system, a user relationship analysis system and an early warning system were constructed on the platform. Through the CMRD, the risk degree of micro-blog net loan contents can be worked out to decide whether warning behaviors are necessary. Through the CMRC, the closeness of net loan bloggers with users can be obtained a propagation map. The key nodes can be cut off by internet security officers in time, which can prevent further propagation of the contents of the microblog net loan. The platform can effectively alleviate the tense situation of college students and illegal online loans.

Keywords: college students' net loan, microblog visualization, risk monitoring, calculation model of relationship closeness, calculation mode of risk degree, propagation map.

1. INTRODUCTION

Net loan is a new industry of Internet finance in recent years. It has the advantages of low interest rate, simple procedure, installments repayment, etc, which is exactly popular among young people. However, it is also easy to be used by illegal merchants [1]. Earlier, some scholars conducted research on bad net loans: Istrate E and others analyzed the extent of bad net loan issues, the causes and the solutions adopted to resolve the bad net loans [2].In recent years, the high frequency new net loan events as "nude loan" and "usury" impact college students badly, which sometimes even threaten their lives [3]. The spread of net loans relies heavily on social media and web advertising. As a new form of social media, microblog Microblogprovides a sharing platform for users to share new things and express new opinions [4]. Therefore, more and more net loan platform and personal loan borrowers in Microblog posted net loan information to attract the attention of college students, but some of the information is with safe risk, and serious consequences may be made without preventing behaviors.

Domestic and foreign scholars have done some researches on visualization and online public opinions. In visualization, Zhao Hua and others made visualization analysis in Microblog users interest identification results [5], Zhao J, et put FluxFlow combined with machine learning in identifying abnormal behavior about popular topics of social media [6], Michael Jendryke and others used the method of Microblog visualization on Sina Weibo to analyze the relationship between news and census [7], Siming Chen and others made visualization analysis of large amounts of data on social media research, summarizes the visual analysis process [8]. In terms of internet public opinion, Xue Gang Chen and others made the network public opinion trend prediction and evaluation research [9], Gu Q and others analyzed the evolution process of network public opinion over time and information source quality evaluation with web crawler tools, which can automatically detect and track hot

Corresponding author: Ruijun Zhang. Email:zrjdoctor@126.com

public opinions [10].

Based on the above theoretical basis, this paper studies a set of early-warning mechanism in the aspect of college students' net loan risk monitoring. In addition, the improvement and innovation is made on Kuo Y H and others' comprehensive analysis method of micro-blog emotion from user's social interaction mode and text view [11], the risk calculation model and relationship intimacy calculation model is made. On the basis of these two models, the net loan risk monitoring system of college students is constructed, which can effectively analyze the risk net loan information and adopt corresponding early-warning behaviors.

2. CORRELATION MODELS

2.1 Calculation model of risk degree

The overall risk value of microblog content can be defined as *DaValue*.

A complete microblog contains emoticons, pictures, links and words, so the calculation model of risk degree needs to be considered from these four aspects.

2.1.1 Building a list of net loan dangerous vocabulary

The list of risk words for net loan contains the following three subsets:

- The high frequency words selected from the examples of microblog, such as *The Next Day's Money*, *Small Private Loan*, *Emergency*, etc. The risk value of the i-th high-frequency vocabulary is set as a_i.
- The names of several major campus loan platforms include aliases, such as *Credit Ease*, *Elite Loans*, *99 Staging*, etc. The risk value of i-th campus loan platform is set as b_i.
- There are common terms of loan, such as Issuing a bid, *Bidding and Freezing a deposit*, etc. The risk value of i-th loan common word is set as c_i.

If there is a risk-word list of net loan (denoted as T_{dan}) in the microblog content as a stored word, then it is set according to the frequency of its occurrence and the risk value of its similar words stored in the list. If its risk value exceeds the lowest value in T_{dan}, then classify it into T_{dan}.

2.1.2 R language processing participle and definition of dangerous values

The Rwordseg program package in R language is used to handle participle words by removing punctuation and dividing the sentences into several words or phrases, and the table function in R language is used for statistic *fr*, which denotes the frequency that each phrase appears. High-frequency words are listed into high-frequency vocabulary in the T_{dan}. In daily language, terms, which are meaningless to the whole sentences, such as *do*, *is* and *have*, are extremely frequent and can be removed from the T_{dan} by computer or hand. The risk value of the name of some campus loan platforms are set in accordance to the development of the platform and its credit. But those frequent words about loan are set value according to their importance in the process of net loan.

As to the expressions and pictures obtained, their risk values can be set correspondingly to their definition combined with its multiplexing values in the field of net loan. Suppose the risk value of the i-th expression is δ_i and the i-th picture is λ_i.

As to the capture link, the final risk value can be figured out by analyzing its content, handling with participles and the risk judging of each word. Suppose the risk of the i-th dark link is π_i.

$$\pi_i = \sum \frac{n_x}{n} \sum \pi_{x,y} \quad (x = a, b, c, \delta, \lambda, \pi) \tag{1}$$

In the formula, n represents the total number of the participle words after segmentations, n_x represents the

number of the words of x , and$\pi_{x,y}$ represents the risk of the y-th word of x type.

2.1.3 Calculation model of risk degree

Combined with the above description, the risk value of microblog content is recorded as:

$$CoValue = \sum X \quad (X = a_i, b_i, c_i, \delta_i, \lambda_i, \pi_i) \tag{2}$$

Besides the impact of content, the risk of microblog net loan is also related with the force of the publishers, which shows specifically on publisher's fans F_c, amount of its likes D_c, amount of its comments L_c and amount of its retweets Z_c [12]. The publisher's influence is recorded as *InValue*.

$$InValue = \sum \lg(x+1) \quad (x = F_c, D_c, L_c, Z_c) \tag{3}$$

Combined with the above description, the calculation model of comprehensive risk degree is obtained as follow:

$$DaValue = InValue \cdot CoValue \tag{4}$$

2.2 Calculation model of relationship closeness

In order to predict the propagation direction of net loan information and understand the relationship between net loan creditors and borrowers, a calculation model of relationship closeness was established on the basis of LDA's microblog user fan intimacy evaluation model [13]. Through analyzing microblog user's relevance, interaction degree and interest similarity, the final calculation model of relationship closeness is ascertained. The model was used to predict the possibility of fans reposting information from bloggers. It should be noticed that the propagation pattern of net loan is called multi-level. The following algorithm is based on the first level of propagation, and closeness values of other levels of propagation can be calculated by this method.

2.2.1 Microblog user's relevant relationship

The relationship between users can be divided into four kinds as follows: Attention (the blogger single concerns), the fans (the uses single concern blogger), mutual fans (the blogger and fans concern mutually) and no relation. If Microblog nodes only establishes the relationship of one-way attention or the fans, it cannot show that there is a mutual trust relationship between both sides, when the two sides reach a certain familiarity forming the friend relationship, interactive feedback has a certain similarity between nodes [14]. Based on this relationship, we believe that the probability of forwarding in these four relationships is from high to low: mutual fans, fans, attention, and no relation. Therefore, the correlation can be simply expressed as equation (5).

The relevance is defined as *Att_degre:*

$$Att\_degree = \begin{cases} 0 & No\ attention \\ 1 & Attention \\ 3 & Fans \\ 4 & Mutual\ \ fans \end{cases} \tag{5}$$

2.2.2 Interaction strength of Microblog users

According to its characteristics, users can use microblog to interact with others by liking, retweeting, @ and commenting. These four methods are also most commonly used. The degree of interaction can be set with these four interactive ways. From the perspective of social capital theory, closeness relationship is based on social interaction [15], the frequency of interaction determines the degree of closeness and the lasting maintenance of the relationship.

In four kinds of interaction, the theoretical research shows that the like is the most simple interaction, forwarding behavior is also relatively simple, and the comment is relatively deep interaction [14], @ was conducted on the basis of the forwarding or comments, it can be simply represented as equation (6).

$$I=\begin{cases}1 & Like\\ 2 & Comment\\ 3 & Retweet\\ 4 & @\end{cases} \tag{6}$$

In order to reflect the interaction between microblog users and fans better, this paper uses bidirectional calculation and weighting method to represent. Define R_{ui} as the interaction degree between the users and his *i-th* fan [13]. Combined with equation （6）, it can be denoted as:

$$R_{ui}=\frac{I}{2}\left(\frac{\sum D_{fi}\cup L_{fi}\cup Z_{fi}\cup A_{fi}+\sum D_{bi}\cup L_{bi}\cup Z_{bi}\cup A_{bi}}{t}\right) \tag{7}$$

In the formula, D_{fi}, L_{fi}, Z_{fi} and A_{fi} denotes the number of the liking, commenting, retweeting and @ of the *i-th* fan f_i for the blogger respectively in a certain period of time t. D_{bi}, L_{bi}, Z_{bi} and A_{bi} denotes the number of liking, commenting, retweeting and @ of the blogger for the *i-th* fan respectively in a certain period of time *t*.

2.2.3 Interest similarity of microblog users

Interest similarity of microblog users can be analyzed by analyzing the topics they participated and the microblogs they retweeted, the final total interest similarity of users will be recorded as S_{bi}.

When a microblog topic is discussed by user, it will be added *# topic name #* and its hyperlinks display in blue words. Each topic that the user participated in will be compared with the topic that the *i-th* fan participates in. If it is the same, then T_{bi} increases by one, where T_{bi} denotes the degree of topic similarity between the blogger and his *i-th* fan in a period of time *t*, and the initial value of T_{bi} is 0.

Users can also forward something according to their own interests. They can attach their own sharing experience, and the bottom of their retweet includes blue words as *@ the original blogger*. By comparing the words, the similarity of interest between users and fans will be reflected further. Through the topic similarity method and contrasting one by one, if it is same then the similarity F_{bi} plus one, where F_{bi} denotes the degree of topic similarity between the blogger and his *i-th* fan in a period of time *t*, and the initial value of F_{bi} is 0.

The final interest similarity of users:

$$S_{bi}=\frac{2\left[\left(B_t-1\right)\frac{T_{bi}}{B_t}+\left(F_t-1\right)\frac{T_{bi}}{F_t}\right]\cdot\sqrt{\frac{1}{B_t}+\frac{1}{F_t}}}{\left(B_t+F_t-2\right)t}+\frac{\left[\left(B_{tr}-1\right)\frac{F_{bi}}{B_{tr}}+\left(F_{tr}-1\right)\frac{F_{bi}}{F_{tr}}\right]\cdot\sqrt{\frac{1}{B_{tr}}+\frac{1}{F_{tr}}}}{\left(B_{tr}+F_{tr}-2\right)t} \tag{8}$$

In the formula, B_t represents the total number of topics that bloggers participated in the discussion in a period of time *t*, F_t represents the total number of fans participated in the discussion of microblog topics in a period of time *t*, B_{tr} represents the total number that bloggers retweeted microblogs in a period of time *t*, F_{tr} represents the total number of microblogs retweeted by fans in a period of time *t*.

2.2.4 Calculation model of relationship closeness

Combining the above three aspects, the calculation model of relationship closeness can be obtained. C_{bi} is set to represent the comprehensive closeness between bloggers and the *i-th* fan:

$$C_{bi}=\prod X \quad \left(X=\mathrm{Re},R_{ui},S_{bi}\right) \tag{9}$$

3. RISK MONITORING PLATFORM OF UNIVERSITY STUDENTS' net loan BASED ON THE MICROBLOG VISUALIZATION

Based on the above calculation model of relational closeness and calculation model of risk degree, a net loan risk monitoring platform of college students is designed, which can be generally divided in visualization system, analysis system of users' relationship and early-warning system. The architecture is shown as Figure 1.

Visualization System: Web crawler crawls cases related with microblog net loan, through the R language and

microblog visualization tools of Peking University (or PKUVIS), and then several propagation relationship diagrams will be obtained. After words segmenting, data filtering and other operations, the system stores the organized data into the database platform. Meanwhile, experts in the field of net loan can add some information that it is hard for the system to identify, such as some sensitive words that has not been recorded into T_{dan} for some time.

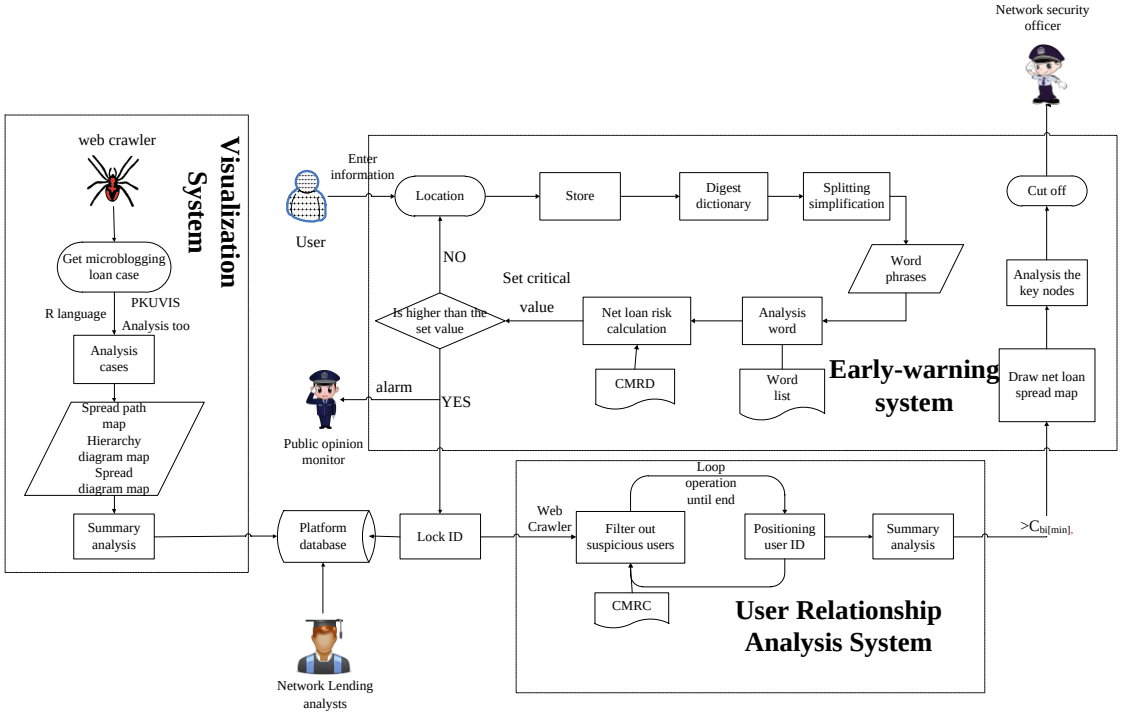

Figure 1. Platform architecture

Early-warning System: the platform define the position of information input by users, split the content sentences [16] and simplify the data, and calculate the overall danger value *DaValue* according to the risk calculation model of T_{dan}. If it belongs to first kind of risk, sign it as , platform will ask public opinion supervisors to make warnings and lock the propagation resource, combined with analysis system of microblog users relationship select which is greater than the minimum value $C_{bi[min],}$, the net loan transmission graph is obtained and the key nodes are analyzed, and then the net security officer is involved to cut it off. If it is the second kind of risk, the platform will ask the user whether to cut off the key node of the transmission graph, and sign it as . If it is the third kind of risk, the platform will continue to pay attention to this information and sign it as .

User Relationship Analysis System: The platform focuses on the high-risk owners' ID and locates the suspicious users of various stages of propagation to summarize and dealt with, through the network crawler and CMRC.

4. DEVELOPMENT OF PLATFORM AND EXPERIMENTAL DATA

4.1 Development technology

The platform uses B/S architecture to develop, SQL Server 2008 is set as the platform of the background database, R language is used to handle the business logic of the middle business layer. In addition, C# language and the .NET Framework are applied as front-end Web development tools, the PKUVIS Analysis tool of Peking University is used as the information display tool of microblog visualization.

4.2 Experimental Data

Firstly select a microblog about net loan. Then the result,which contains the experimental data, propagation route map and the distributed map of propagation information ,will be obtained by functional testing through the platform. The core of processing data in the platform is parallel data mining because the Internet Information is complicated and gigantic [17].

For example, there is the microblog news about net loan that has been forwarded for more than 2000 times. The original content is as [Xiamen net loan suicide girl's last week: under crazy struggle, she considered suicide for more than once]. No one knows Xueqi (a pseudonym) experienced what kind of psychological struggle in a small hotel near a university in east of Quanzhou. On the early morning of April 11, she chose to end her life and the police initially regard it as suicide by burning charcoal. On the 14th, the paper contacted with two of her closest friends who told reporters: https://wx4.sinaimg.cn/large/005vnhZYgy1feontjyg1rj30fq0f20x5.jpg. To reduce redundancy, a propagation chain is taken as an example.

Set $C_{bi[min]} = 0.00020$,the critical value of the first kind of risk is 230.0, the critical value of the scend kind of risk is 100.0

Table1. The information chart of original microblog data

No.	User name	Retweetings	Likes	Comments	Fans	Number of microblogs	Concerns
1	Pengpai News	6137.00	22440.00	23848.00	5604398.00	28434.00	310.00

Table2. The information chart after the split of original microblog

No.	High frequency word	Risk value	High frequency word	Risk value	High frequency word	Risk value
1	Net loan	4.68	Commit suicide	5.78	Myself	0.78
2	Suicide	5.88	No	1.23	Life	2.65
3	Girl	2.12	College	3.24	Police	3.33
4	Last	3.14	Hotel	1.87	First step	1.24
5	A week	4.33	Experience	1.67	Identification	2.66
6	Crazy	3.12	Psychlogy	1.21	Choice	0.54
7	Dun	4.55	Struggle	2.45	End	4.21
8	More	1.23	Early morning	1.22	Thought	0.54
9	Once	0.78				

According to the content of this microblog and combining with CMRD:

Risk value of high frequency word: $a = 4.68 + 5.88*3 + ... + 0.54 = 76.21$;

Risk value of platform word: $b = 0$;

Risk value of loan words $c = 0$;

Risk value of expressions $\delta = 0$;

Risk value of pictures $\gamma = 2.33$;

Risk value of dark links $\pi = 0$;

Risk value of content information is obtained with equation(2): $CoValue = 78.54$

Risk value of publishers' force is obtained with equation (3): $InValue = 19.27$

Risk value of final information is obtained with equation(5): $DaValue = 1513.4658 \in [230, +\infty)$ It is the first kind of risk and it is necessary to obtain the propagation graph according to CMRC

Table3. The required information chart of CMRC after handling

Number	Dfi/t	Dbi/t	Lfi/t	Lbi/t	Zfi/t	Zbi/t	Afi/t	Abi/t	Re
1	7.00	-	3.00	-	0.00	-	0.00	-	-
2	0.00	2.00	1.00	4.00	0.00	2.00	0.00	0.00	1.00
3	5.00	1.00	1.00	2.00	0.00	2.00	0.00	0.00	1.00
4	1.00	3.00	1.00	4.00	1.00	1.00	0.00	0.00	1.00
5	0.00	0.00	2.00	2.00	0.00	1.00	0.00	0.00	1.00
6	-	1.00	-	0.00	-	1.00	-	0.00	3.00

In the table, $t = 48h$.

The degree of interaction and interest similarity between users at all levels is obtained with equation (7), (8) and Table 3.

$R_{1,2} \approx 0.35$ $R_{2,3} \approx 0.15$, $R_{3,4} \approx 0.26$, $R_{4,5} \approx 0.15$, $R_{5,6} \approx$

$S_{1,2} \approx 0.016$ $S_{2,3} \approx 0.008$, $S_{3,4} \approx 0.003$, $S_{4,5} \approx 0.026$, $S_{5,6} \approx 0.010$

The ultimate degree of closeness at all levels is obtained with equation (9)

$C_{1,2} = 0.00560$, $C_{2,3} = 0.00120$, $C_{3,4} = 0.00078$, $C_{4,5} = 0.00390$, $C_{5,6} = 0.00270$

They are all more than or as same as $C_{bi[min]} = 0.00020$, and a propagation graph in the red box of figure 2 is obtained.

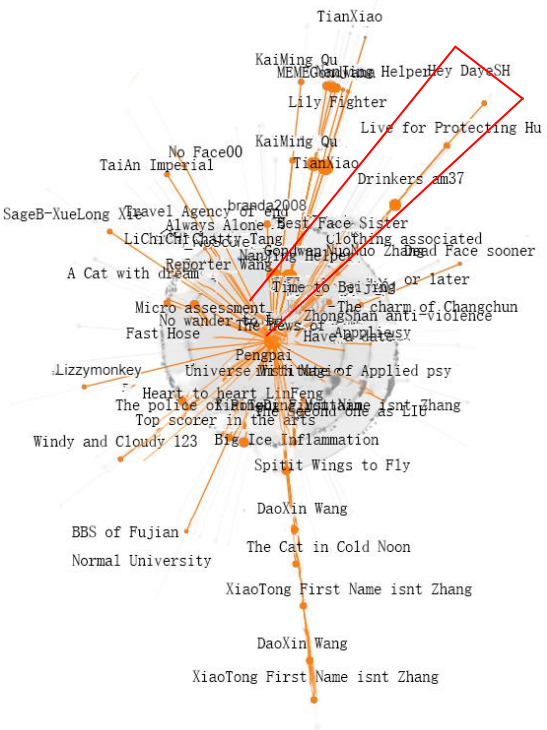

Figure 2. Propagation Route Map

In Figure 2, each circular node represents a user, and the others retweet and comment around the central microblog. The bigger the node is, the greater the impact of this user will be. By combining the propagation route map with those second-retweet users IDs, it can be observed more distinctly and directly.

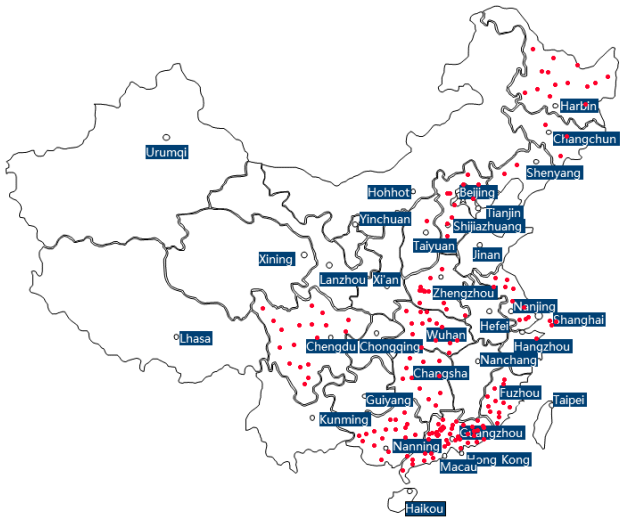

Figure 3. Distributed Map

According to the information of selected microblog, the distributed map can be drawn with population distribution. Each point represents each blogger who retweeted. In the same region, the intensity of the points is proportional to the amount of net loan information forwarding. From Figure 3, it is clear that the retweet frequency of this microblog net loan reaches highest in GuangDong Province. Therefore, its network supervision department should focus on it and take targeted programs.

5. CONCLUSION

The article designed a net loan risk monitoring platform of college students based on the microblog visualization. The platform combines calculation model of intimacy and risk degree, it can analyze and predict the transmission direction of net loan information in time, and draw the net loan transmission map. By cutting off the key nodes, the internet security officers will stop the illegal net loan information from the source and avoid the occurrence of malignant events. The platform takes advance warning mechanism of microblog contents, which can effectively prevent illegal net loan information through micorblog into the field of college students, and provide supervision departments convenient tools to control the spread of illegal net loans. The methods studied in this paper can purify the environment of microblog to some extends, reduce the probability of college students falling into illegal net loans and protect the safety of life and property of college students.

ACKNOWLEDGEMENTS

This research is supported by the 2017 National Undergraduate Innovative Training Project - An Net Loan Monitoring Platform for University Students Based on Visual Micro-blog, under the Grant NO. 201710488005 , by the Humanities and Social Sciences Research Project from the Hubei Provincial Department of Education, under the Grant NO.18D014, by the Service Science and Engineering Research Project from Wuhan University of Science and Technology ,under the Grant NO.CSSE2017GA07.

REFERENCES

[1] Cai Yu.(2016). Current situation and strategy of college students' online loan [J]. Modern economic information, (15):309+311(in Chinese).

[2] Istrate E, Das Gupta D, Weissburg P. (2007). Toward Developing a Structured Approach to the Diagnosis and Resolution of Nonperforming Loans: the Case of China and India[J]. Review of Policy Research, 24(4):345-365.

[3] Zhang Jie, Mu Shu. (2017). Analysis of college students' online loan cognition and its risk prevention -- based on the survey data of tianjin university [J]. Credit, 35(4):22-26(in Chinese).

[4] Zhao Y, Qin B, Liu T, et al.(2016). Social sentiment sensor: a visualization system for topic detection and topic sentiment analysis on microblog[J]. Multimedia Tools & Applications, 75(15):8843-8860.

[5] Zhao Hua, Ji Xiaowen, Zeng Qing-tian, et al. (2015). Interest identification and visualization of microblog users based on topic related space [J]. Computer science, 42(S1):500-502+509 (in Chinese).

[6] Zhao J, Cao N, Wen Z, et al. (2014). FluxFlow: Visual Analysis of Anomalous Information Spreading on Social Media[J]. IEEE Transactions on Visualization & Computer Graphics, 20(12):1773-1782.

[7] Michael Jendryke,Timo Balz,Mingsheng Liao.(2017). Big location‐based social media messages from China's Sina Microblog network: Collection, storage, visualization, and potential ways of analysis[J]. Transactions in GIS, 21(4):825-834.

[8] Chen S, Lin L, Yuan X.(2017). Social Media Visual Analytics[J]. Computer Graphics Forum, 36(3):563-587.

[9] Chen X G, Duan S, Wang L. (2017). Research on trend prediction and evaluation of network public opinion[J]. Concurrency and Computation: Practice and Experience, 29(24): 4212-4218.

[10]Gu Q, He X, Wang x. (2014). Study on evolution trends of network public opinion based on hyperlink analysis[J]. Journal of Digital Information Management, 12(6):421-428.

[11] Kuo Y H, Fu M H, Tsai W H, et al.(2016). Integrated microblog sentiment analysis from users' social interaction patterns and textual criticism [J]. Applied Intelligence,44(2):399-413.

[12] Lliu Gongshen, Meng Kui, Xie Jing.(2014). An early warning algorithm for microblog [J]. Computer science, 41(12):33-37(in Chinese).

[13] Wang Qiuxun, Yu Haoliang, Xu Haocheng, etc. (2016). Based on LDA's microblog user fan affinity evaluation model [J]. Computer application and software, 33(10):67-71(in Chinese).

[14] Xiong Jianying. (2017). A study on the model of emotional anomaly of microblog users based on trusted feedback [J]. Intelligence science, (4):48-53(in Chinese).

[15] Yao Zhengyu.(2013). Research on Intimacy and Social Network Interaction Behavior[C].//Communication and China Fudan Forum. Networked Relationship: New Communication and Current Chinese Proceedings, 1(in Chinese).

[16] Wubben S, Antal V D, B Krahmer e. (2012). Sentence Simplification by Monolingual Machine Translation [C] / / Proceedings of the Annual Meeting of the Association for Computational Linguistics, 1015-1024.

[17] Ni Yingjie, Wang Yuke, Zhang Jun. (2009). Network mass information processing platform based on high-performance data mining [J]. Computer engineering and science, 31(S1):129-132(in Chinese).