

Public Opinion Detection in an Online Lending Forum: Sentiment Analysis and Data Visualization

Ge Zhan

Division of Business & Management
Beijing Normal University-Hong Kong Baptist University
United International College (UIC)
Zhuhai, P.R. China
e-mail: garygezhan@uic.edu.hk

Ming Wang

College of Economics
Shenzhen University
Shenzhen, P.R. China
e-mail: wmtroy@outlook.com

Meiyi Zhan

School of Computer and Information Engineering
Beijing Technology and Business University
Beijing, P.R. China
e-mail: 727027376@qq.com

Abstract—Public opinion entails emotions, wishes, attitudes and ideas among individuals and various social groups in a certain historical stage and social space. This research aims to employ natural language processing (NLP), sentiment analysis and data mining technologies to build a public opinion analysis system to serve enterprises' need of online public opinion detection. By using the data set from an online peer-to-peer (P2P) lending forum with 4148 reviews, we analyzed the data with data visualization techniques.

Keywords—online public opinion; data visualization; sentiment analysis; data mining; opinion detection

I. INTRODUCTION

In the era of online big data, the continuous development of mobile network technology enables the application of new media dramatically. Individuals can efficiently spread online information through social networking sites (SNS) [1-2]. When a growing number of users are interested in social events, phenomena, comments and supervision, the positive or negative effects of online information sharing are considerably amplified.

However, the vast majority of enterprises do not have effective means to identify and deal with the negative information which may damage their reputation and brand value, and information about product market, consumer attitude, industry development and competitive intelligence, in a timely and accurate manner [3]. At present, many enterprises adopt traditional manual monitoring methods to approach public opinion. Artificial monitoring is insufficient to meet the needs of the enterprises. Public opinion related to the focal enterprise is spread widely on the Internet, but it cannot be easily collected and managed accurately. This research aims to employ natural language processing (NLP), sentiment analysis and data mining technologies to build a public opinion analysis system to serve enterprises' need of online public opinion detection.

II. PUBLIC OPINION DETECTION FOR ENTERPRISES

Public opinion is the sum of multiple emotions, wishes, attitudes and ideas among individuals and various social groups in a certain historical stage and social space. Internet public opinion can be developed from online speech, and is a specific form of public opinion. Whether Internet speech can develop into Internet public opinion largely depends on whether the topic of Internet speech has strong activeness and sensitivity. Online public opinion is diversified, spread quickly and interactive, which has the incomparable advantages over traditional media. Online public opinion has positive and healthy aspects, as well as negative and decadent aspects [4-5].

The enterprise public opinion is limited to the public opinion content related to a certain enterprise subject. It is a collection of a variety of emotions, attitudes and opinions expressed through the Internet by the vast number of Internet users in a specific period for certain product production or service provided by a specific enterprise. With greater use of online media, enterprises pay more and more attention to the impact of their online public opinion on their corporate image. Network blog, SNS (brand forum, wechat, QQ, etc.), and e-commerce platform are the main channels of online public opinion communication. The enterprise network public opinion has the general characteristics of network public opinion: the virtual space of public opinion information, the real-time nature of public opinion communication, the interactivity of opinion published, disseminated and fed back by online participants, and the openness of online opinion communication [6].

Online public opinion management concerns enterprises' own interests on the network. When obtaining the original information from the public opinion information source, the information can be filtered pertinently, which greatly reduces the information processing capacity of the enterprise public opinion management system. The accumulation of corpus and the interaction history, and the interaction between

enterprises and customers, tend to concentrate on official micro-blog, Twitter, product and industry forums, so it is easy to accumulate corpus [7]. However, government public sentiment management often lacks sufficient corpus.

The network public opinion of enterprises often spreads and breaks out in a large scale in a short period of time due to an emergency, and its impact will be immediately transmitted to the production and sales performance of enterprises. Therefore, the timeliness of monitoring and management of network public opinion of enterprises is relatively high. However, the government's management of social public opinion involves a larger scope, the time for correct guidance of online public opinion is relatively long, and the influence of public opinion is more average [8].

The public opinion management of the enterprise has strong target, mainly for the relevant groups of the enterprise, including customers, partners, main competitors, potential target customers, etc. [9-10]. The government level social public opinion management aims at the whole country's all-round public opinion management, which is broader and less targeted compared with enterprises. The management of public opinion in enterprises tends to be actively used and exerted, while the management of public opinion at the government level tends to be supervised and prevented. Specifically, corporate public opinion is generally to collect, summarize, analyze and further spread the information of corporate brand reputation, competitor dynamics, industry status and hot events, so as to achieve the purpose of commercial profit [11]. While the government level social public opinion management focuses on the use of the network to timely discover the information that has an adverse impact on national interests and timely supervise and handle, so as to minimize the adverse impact on society [12].

III. METHOD

We collected online post and review data from forum.lendacademy.com, which is among the largest P2P lending sites in U.S. Overall, 4148 reviews of 300 posts were collected. The data cover over two years from 2017 January to 2019 October.

A SaaS API for sentiment analysis was employed to conduct the sentiment analysis. The sentiment polarity of the posts is measured with a three-way classification: positive, neutral and negative sentiment, coded as 3, 2 and 1 respectively.

The sentiment scores were estimated with a confidence value, where higher confidence value with a positive sentiment indicates greater confidence with the polarity. The confidence value ranges from 0 to 1, where 1 denotes the highest possible confidence. The sentiment score equals to the sum of sentiment level (1, 2 and 3) and confidence value (0~1). The sentiment scores in the sample range from 1.388 to 3.929 with negative polarity from 1.388 to 1.999, neutral polarity from 2.375 to 2.973, and positive polarity from 3.370 to 3.929.

IV. RESULTS

Internet public opinion events have a unique life cycle, which refers to the changes of Internet public opinion in the

whole process from occurrence to end. According to the change dimension of internet users' emotions, the development of public opinion can be divided into incubation period, growth period, outbreak period, and recession period [13-14]. Figure 1 shows the change with sentiment scores for the posts in lendacademy.com from 2017 January to 2019 October. Negative posts were more frequently posted than positive posts, particularly in 2019.

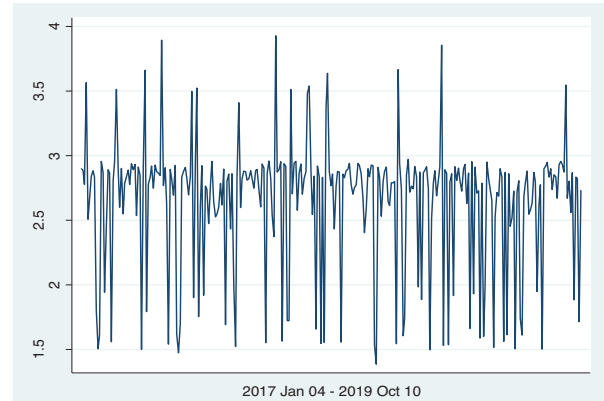


Figure 1. Time Line of Sentiment Scores.

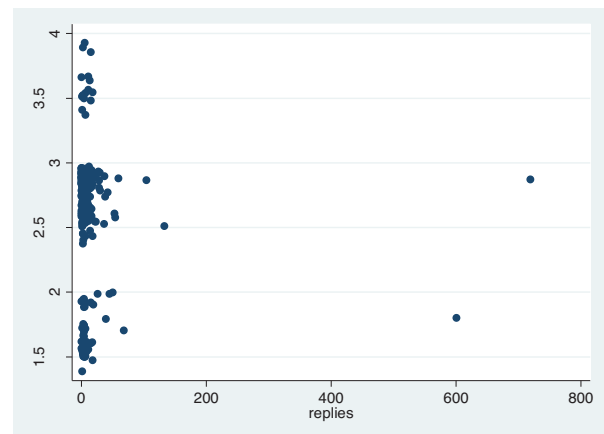


Figure 2. Sentiment scores with number of replies.

Data visualization technology can illustrate large-scale data in an intuitive visual way. Therefore users can make full use of their capacity to observe and process information, so as to find the relationship between information and hidden patterns [15]. Based on the sentiment analysis of data from an online P2P lending forum, this paper analyzed the data with scatter charts (see Figure 2 and 3).

V. DISCUSSION AND CONCLUSION

As indicated by the data visualization above, we can see negative events gain more views and replies. Once the events are exposed, users quickly responded to the news and entered the outbreak period. On the one hand, social media information based on Web 2.0 technology has a fast speed, wide range and great influence. Internet users have a high degree of attention to social ethics, moral standards and other issues, and the enthusiasm to communicate their opinion

online [16]. When sensitive events occur in the forum, it is easy to trigger public discussion [17-19]. Time line is an effective way to show the evolution of data in time dimension. It connects one or more events in series according to time order, forms a relatively complete record system, and then presents it in the form of graphics and text.

The findings from this study indicate that the higher confidence value with the negative posts, the more likely the post was viewed and replied. Online platform managers, therefore, should frame the negative news with a clear signal. Users are more likely to share with others when the message has a clear negative polarity. Positive posts with higher confidence values, on the other hand, did not generate more replies but they were more likely viewed. While most previous studies suggest that negative information is more effective for promotion purpose [20-21], our findings shed interesting new lights in that negative posts in an online P2P setting are not driving more views nor replies compared with posts with neutral sentiment.

Public opinion events have a unique life cycle. The life cycle of online public opinion refers covers the whole process from the occurrence to the end of the opinion events. From the visualized data shown above, it can be seen that the fermentation period of negative sentiment becomes relatively short in 2019. From one negative-sentiment event to the next, it was roughly half a month. Once the event was exposed, it would quickly be spread.

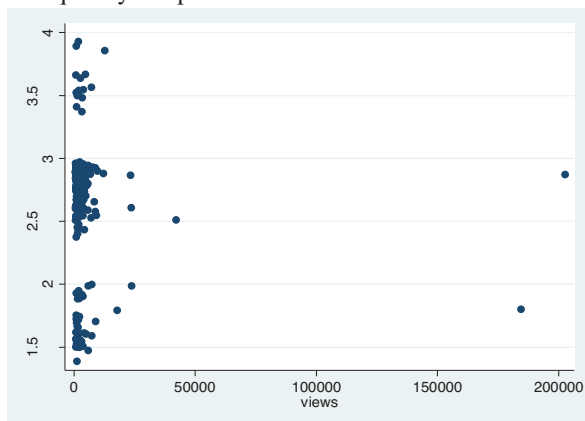


Figure 3. Sentiment scores with number of views.

REFERENCES

- [1] J. Tang, J. Wang, and C. Wu, "Research progress on network public opinion based on rough sets from the big data perspective," *Proc. IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, IEEE, May 2019, pp. 1074-1077.
- [2] S. Peng, G. Wang, and D. Xie, "Social influence analysis in social networking big data: Opportunities and challenges," *IEEE network*, vol. 31, 2016, pp. 11-17.
- [3] L. Bing and K. C. Chan, "A fuzzy logic approach for opinion mining on large scale twitter data," *Proc. IEEE/ACM 7th International Conference on Utility and Cloud Computing*, IEEE, December 2014, pp. 652-657.
- [4] D. Jiang, X. Luo, J. Xuan, and Z. Xu, "Sentiment computing for the news event based on the social media big data," *IEEE Access*, vol. 5, 2016, pp. 2373-2382.
- [5] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent systems*, vol. 28, 2013, pp. 15-21.
- [6] Q. Li, Y. Chen, J. Wang, Y. Chen, and H. Chen, "Web media and stock markets: A survey and future directions from a big data perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, 2017, pp. 381-399.
- [7] X. Fang and J. Zhan, "Sentiment analysis using product review data," *Journal of Big Data*, vol. 2, 2015, p.5.
- [8] S. A. Munson and P. Resnick, "Presenting diverse political opinions: how and how much," *Proc. SIGCHI conference on human factors in computing systems*, ACM, April 2010, pp. 1457-1466.
- [9] A. Gupta and P. Kumaraguru, "Credibility ranking of tweets during high impact events," *Proc. The 1st workshop on privacy and security in online social media*, ACM, April 2012, p. 2.
- [10] M. J. Lee and J. W. Chun, "Reading others' comments and public opinion poll results on social media: Social judgment and spiral of empowerment," *Computers in Human Behavior*, vol. 65, 2016, pp. 479-487.
- [11] P. Sobkowicz, M. Kaschesky, and G. Bouchard, "Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web," *Government Information Quarterly*, vol. 29, 2012, pp. 470-479.
- [12] H. Chen and D. Zimbra, "AI and opinion mining," *IEEE Intelligent Systems*, vol. 25, 2010, pp. 74-80.
- [13] X. Liu, L. Xu, and J. Zhao, "Co-extracting opinion targets and opinion words from online reviews based on the word alignment model," *IEEE Transactions on knowledge and data engineering*, vol. 27, 2014, pp. 636-650.
- [14] C. Kaiser and F. Bodendorf, "Opinion and relationship mining in online forums," *Proc. IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 1, IEEE, September 2009, pp. 128-131.
- [15] D. Shen, C. Krumme, and A. Lippman, "Follow the profit or the herd? Exploring social effects in peer-to-peer lending," *Proc. IEEE Second International Conference on Social Computing*, IEEE, August 2010, pp. 137-144.
- [16] P. H. Calais Guerra, A. Veloso, W. Meira Jr, and V. Almeida, "From bias to opinion: a transfer-learning approach to real-time sentiment analysis," *Proc. The 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, August 2011, pp. 150-158.
- [17] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: quantifying influence on twitter," *Proc. The Fourth ACM International Conference on Web Search and Data Mining* February 2011, pp. 65-74.
- [18] R. Vedala and B. R. Kumar, B.R., "An application of naive bayes classification for credit scoring in e-lending platform," *Proc. International Conference on Data Science & Engineering (ICDSE)*, IEEE, July 2012, pp. 81-84.
- [19] H. Zhao, Y. Ge, Q. Liu, G. Wang, E. Chen, and H. Zhang, "P2P lending survey: platforms, recent advances and prospects," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, 2017, pp. 72.
- [20] C. M. Cheung and D. R. Thadani, "The impact of electronic word-of-mouth communication: A literature analysis and integrative model," *Decision support systems*, vol. 54, 2012, pp. 461-470.
- [21] D. Yin, S. Mitra, and H. Zhang, "Research note—When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth," *Information Systems Research*, vol. 27, 2012, pp. 131-144.