Research Progress of Network Public Opinion Based on Fuzzy Set from the Perspective of Big Data

Jianguo Tang*, Jianghua Wang† and Fang Li‡

1 School of Artificial Intelligence and Big Data, Chongqing College of Electronic Engineering, Chongqing, China
2 School of Computer Engineering, Changji University, Changji, China
Email: tjguo@126.com

Abstract. Granular computing is a new method of intelligent information processing. It describes knowledge and simulates human thinking mode in a granular way. It can realize the rapid transformation between different granular knowledge and is suitable for the analysis and modeling of large-scale and complex data. Fuzzy set is an important model in granular computing, which describes uncertain problems through fuzzy membership function and has been widely used in many fields such as network public opinion analysis. This paper combs the application research of fuzzy sets in network public opinion analysis from the perspective of big data, and discusses the different roles of fuzzy sets in network public opinion analysis from four aspects: fuzzy comprehensive evaluation, fuzzy reasoning, fuzzy granularity and generalized fuzzy set. The characteristics of different methods are compared and the unique features of fuzzy sets in these four aspects are pointed out. The unique advantages and some problems are discussed as well, and the future trend of this field is discussed.

1. Introduction
Nowadays, with the continuous innovation of information technology and the deep integration of digital technology and network technology, big data has become an effective tool to reflect public opinion and will in the new era, and gives a new decision-making way for the research of network public opinion. Big data is characterized by some “V” like Volume, Velocity, Variety and Value, which bring many challenges to the existing public opinion analysis models. The fuzzy set model is viewed as an effective way to process uncertain problems. It simulates the human thinking way to solve problems to process data and provides an effective paradigm for processing large-scale and complex data analysis in the field of artificial intelligence [1]. The computing framework of big data mining is highly consistent with the computing paradigm contained in granular computing, which enables granular computing to meet the challenges of network public opinion analysis in big data era.

2. Granular Computing and Fuzzy Set
Granular computing, originally called information granule, was proposed firstly by Zadeh [2-3]. Lin [4] conducted research on granular computing under the neighbourhood relationship. On this basis, Yao [5-6] uses the granular computing rough set theory for knowledge mining, which is applied to rule extraction and intelligent data processing. The existing granular computing research includes single granularity and multi-granularity. Single granularity research is generally used to deal with uncertainty [7-10]; multi granularity research [11-14] is to better adapt to the cognitive ability of human multi-level thinking, in order to effectively solve complex problems.
Fuzzy set is one of the common methods of granular computing, which was proposed by Zadeh [15] in 1965. It uses prior knowledge to give a vague description of things. In specific modelling, the prior knowledge is expressed as a membership function, and the degree of similarity or closeness between different things is measured according to the membership degree determined by the membership function.

Let $U$ be a universe of discussion, $x \in U$. $\mu_\delta$ is a map on $U$ and $\mu_\delta : U \rightarrow [0,1]$, $x \mapsto \mu_\delta(x)$. Then $\mu_\delta$ can induce a fuzzy set $A$ on $U$, and $\mu_\delta$ is the membership function of $A$ and $\mu_\delta(x)$ is the grade of membership of $x$ with respect to $\delta$.

The fuzzy set model now has been used to analyse the network public opinion. Scholars have carried out a series of studies using fuzzy set methods, such as online public opinion early-warning, online public opinion threat estimation, online public opinion evolution and online public opinion transmission, etc.

3. Application of Fuzzy Set Model in Network Public Opinion
Because of the fuzziness and uncertainty of the information need to be processed in network public opinion analysis, fuzzy set method is just a powerful tool to solve this problem and has been used to analyze the network public opinion. The fuzzy comprehensive evaluation method can effectively evaluate the public opinion affected by various factors in the fuzzy environment. The fuzzy reasoning method can simulate human thinking to complete decision-making and carry out public opinion early-warning and judgment. Fuzzy granular model can simulate the process of formation and evolution of network public opinion.

3.1. Fuzzy Comprehensive Evaluation
The fuzzy comprehensive evaluation method uses the synthetic mechanism of fuzzy relations to analyse several influencing factors that belong to the grade of the evaluated object. Generally, it includes four steps: determining factor set, determining evaluation set, calculating index weight and fuzzy comprehensive evaluation. Now it has been widely used in the construction of public opinion analysis index system. Shi [16] used the dynamic fuzzy set and fuzzy comprehensive evaluation to build the public opinion early-warning index system, and selected two categories of public opinion audience and public opinion, a total of seven indicators for the comprehensive evaluation and dynamic early-warning of public opinion on the network. Wang [17] built an index system including four categories of indicators: emergencies, public opinion netizens, public opinion media and public opinion situation. The mixed weighting method is used to calculate the weights, and the fuzzy operator is selected to obtain the numerical evaluation value of each public opinion index, so as to obtain the warning level. Wang [17] further combined the BP neural network [18] and Elman neural network [19] to build a network public opinion threat estimation model based on the fuzzy comprehensive evaluation method. The public opinion data is processed automatically by computer, and the simulation experiment is carried out. Zhang [20] based on fuzzy comprehensive evaluation to construct an index system consisting of three types of indicators: information flow generation risk, information flow transmission and diffusion risk, and information flow recovery risk, and conducts an empirical evaluation study on the risk of public opinion information flow in emergency network. Li [21], Zhang [22], and Dai [23] and other scholars constructed a multi-layer fuzzy comprehensive evaluation model to carry out network public opinion security situation assessment, network public opinion classification research and early-warning, and network public opinion security evaluation research. Wang [24] combines AHP and fuzzy comprehensive analysis to build the public opinion early-warning model of AHP fuzzy comprehensive analysis mobile social network. Chen [25] uses the analytic hierarchy process (AHP) to construct a risk early-warning model of social network public opinion based on ANP grey fuzzy. The network analytic hierarchy process is an optimization method of the analytic hierarchy process, which is more suitable for dealing with the complex relationship of mutual transformation and mutual influence of the influence factors of network public opinion, and can better determine the weight of the early-warning indicators of network public opinion risk.
3.2. Fuzzy Reasoning
Fuzzy reasoning refers to the process of reasoning through fuzzy sets and relations to reach a conclusion without accurate value. Because the human reasoning process is usually approximate, some scholars introduce fuzzy reasoning logic into public opinion analysis and use fuzzy reasoning to make decisions. Li [26] uses the intuitionistic fuzzy set theory proposed by Atanassov [27-28] to select the public opinion analysis model which is convenient for computer to realize automatically, and complete the judgment of public opinion early-warning level by computer instead of manual, and verifies the validity of the method through experiments. Lin [29] further normalized the five analysis indicators of topic importance, sentiment tendency, topic attention, number of the public, and speed of transmission, using Gaussian functions to find their membership, and finally using fuzzy reasoning to automatically judge public opinion using a computer early-warning level. Wang [30] designed a fuzzy reasoning algorithm to find the maximum fit between the intuitionistic fuzzy subset of public opinion threat level and the reasoning result, and then determined the threat level. Zhang [31] transformed the fuzzy hierarchy matrix into the fuzzy consistency matrix to judge the weight of index elements, then the intuitionistic fuzzy reasoning is used to carry out the monitoring and early-warning evaluation of network public opinion, and the intuitionistic fuzzy theory is optimized. Based on the membership relationship, fuzzy public reasoning is used to determine the level of Internet public opinion early-warning. Xiao [32] carried out research on public opinion early-warning levels of Weibo based on intuitionistic fuzzy reasoning, and used Gauss type in fuzzy mathematics to determine the weight of each attribute in the index system through the intuitionistic fuzzy entropy of similarity.

3.3. Fuzzy Granularity
Cellular automata is a dynamic model, which is discrete in time and space. It is an effective tool to explore and simulate nonlinear complex systems. In recent years, it has been gradually applied to network public opinion analysis.

Chen [33] combined fuzzy reasoning and cellular automata to propose an evolution model of network public opinion, and analysed and studied the evolution process of network public opinion. Zhang [34] designed a cellular automata model of Internet public opinion communication based on fuzzy reasoning, and analysed the problems of "clustering" and "unification" in public opinion communication by adding the parameters of emotional inclination. Dang [35] combined fuzzy reasoning and granular theory, designed two input variables of environmental fitness and preference, and built a fuzzy granular model of public opinion transmission. It was concluded that after communication and discussion, the group would form a discount Comments. Dai [36] introduced fuzzy reasoning into the network public opinion communication cellular machine, studied the phenomenon of "group polarization" in the process of public opinion communication, and simulated and analysed its formation and evolution law. Based on the adaptive neuro fuzzy reasoning and cellular machine, Yu [37] designed three variables, that is the stubbornness, understanding of events and influence of friends, and the Wechat users are took as the object to predict the probability of group attitude change to public opinion events.

3.4. Other Fuzzy Sets
In addition to the above three common fuzzy set methods, scholars also use fuzzy integral, triangular fuzzy number, fuzzy information entropy, fuzzy data envelopment analysis and fuzzy logic to analyse the public opinion of big data network.

Wang [38] based on Choquet fuzzy integrals, researched the Internet public sentiment around the four attributes of propensity, attention, sensitivity and focus, and conducted an empirical analysis on the tightness of the interconnection between topics and the public. Liu [39] will combine the Fuzzy Delphi method based on the triangular fuzzy number with the fuzzy analytic hierarchy process, screen indicators and calculate weights, and build a public opinion early-warning indicator system. Zhang [40] used interval fuzzy information to calculate the index weights, and then calculated the evaluation value of alternatives through weighted aggregation operation, to assist decision makers in selecting the optimal scheme from the set of alternative emergency decision schemes. Wang [41] uses fuzzy logic to perform
user and event sentiment calculation, and then, uses fuzzy clustering and fuzzy pattern mining to analyse and analyse sentiment communities in Weibo. Liu [42] based on fuzzy theory and used the DEA model method to screen the evaluation index of the effect of microblog information dissemination, using the back-off method to calculate the relative efficiency vector of each evaluation unit when the input variables or output variables are eliminated. Use Euclidean distance to construct evaluation index system. Hu [43] uses the fuzzy neural network method which integrates the fuzzy theory into the neural network to build the trend prediction model of microblog public opinion, and optimize the parameters of the fuzzy neural network to improve the accuracy of public opinion prediction. Fan [44] constructed an emergency group decision model based on interval-valued fuzzy entropy, used interval-valued fuzzy entropy to calculate weights, and evaluated network public opinion emergencies.

4. Conclusion and Prospect
In recent years, many research works about network public opinion based on the fuzzy set have been published and more and more researchers put attention to this research filed. However, these studies still have many shortcomings that they not yet formed an analysis system of complete network public opinion. A complete network public opinion analysis system, from the perspective of big data, needs to include three main modules: collection of network public opinion data, analysis model of network public opinion, and early-warning model of network public opinion. The first module considers storing massive amounts of data with Hardoop cloud storage technology, and collecting all kinds of network public opinion data with a new web crawler technology. The second module is to construct an analysis model of public opinion to generate a decision rule set. The third module will use the rough set model to construct an early-warning model of network public opinion, which will obtain public opinion response strategies based on the expert knowledge, and use actual situations and test results to improve the model.

The rough sets is another granular computing model to solve uncertainty problems. Unlike rough sets, in fuzzy sets, the membership function is constructed based on the experience of expert, which is subjective and lack objectivity of rough sets. Nevertheless, in rough sets, the value of membership function can be obtained directly from the data which need to be processed. In other words, rough set is a method without subjective experience. This shows that rough set and fuzzy set have good complementarity and correlation. Therefore, we can combine them to construct new data processing models such as fuzzy rough sets or rough fuzzy sets to enrich the connotation of granular computing theory. In our future study, the advantages of these models can be applied to solve problems in field of network public opinion. Firstly, the attribute reduction theory of rough set is used to mine new valuable indexes from public opinion data, and the weight of each index is determined by attribute importance method. Secondly, the generalized rough fuzzy set method is used to describe the indicators in a fuzzy way, and the degree of coupling between the indicators is evaluated using the closeness method of the fuzzy set to achieve a reasonable organization of the indicator set. Then, on the basis of the existing data granulation methods, combined with the diversity characteristics of network public opinion data to develop the method of data diversification granulation, and then design the diversified granularity structure of data. Then, the decision set such as the three-branch decision in granular computing is used to create a rule set, and then the rule set is granulated to build a network public opinion monitoring model. Finally, we use the methods of information entropy and attribute importance to select the early-warning index and set the index weight, and use the methods of variable precision rough set, and combine the factors such as real demand to set the early-warning threshold, and then build the early-warning model of network public opinion.

As a powerful tool to deal with uncertain problems, fuzzy sets have attracted the research interest of many scholars. They applied fuzzy sets to discuss network public opinion issues and obtained many valuable results. For better understand the positive role of fuzzy sets in online public opinion study based on big data, this article summarizes the related research results, and points out the advantages and disadvantages of these studies, and provides some suggestions to solve these problems. We hope this work can provide people with a good reference and help.
Acknowledgments
This work was partially supported by the National Natural Science Foundation of China under Grant No. 61562079, and Youth Project of Science and Technology Research Program of Chongqing Education Commission of China under Grant No. KJQN201903113, and Xinjiang Uygur Autonomous Region General Higher Education Reform Research Project under Grant No. 2017G084.

References
[1] Liang J, Qian Y, Li D and Hu Q 2015 China Science 11 1355.
[2] Zadeh L 1979 Advances in Fuzzy Set Theory and Applications (Amsterdam: North-Holland) p 3.
[3] Zadeh L 1997 Fuzzy Sets Syst. 90 111.
[4] Lin T Y 1999 Computing with Words in Information/Intelligent Systems (New York: Springer Verlag) pp 183-200.
[5] Yao Y 2007 Proc. IEEE/WIC/ACM Int. Conf. on Web Intelligence (United states: Institute of Electrical and Electronics Engineers Computer Society) p xxxi.
[6] Yao Y 2013 Lecture Notes in Computer Science (Canada: Springer Verlag) p 16.
[7] Miao D, Xu F, Yao Y and Wei L 2012 Chinese J. Comput. 2 2351.
[8] Wang G, Zhang Q, Ma X and Yang Q 2011 J. Software 4 676.
[9] Wu W 2009 IEEE Int. Conf. on Granular Computing (China: IEEE Computer Society) p 596.
[10] Liu Q, Sun H and Wang H 2008 Chinese J. Comput. 4 543.
[11] Lin G, Liang J, Qian Y 2014 J. Fronti. Comput. Sci. Techno. 2 45.
[12] Liu C and Miao D 2013 Inform. Techno. J. 12 506.
[13] Bao W, Hu Q and Wang C 2013 J. Nanjing University (Nat Sci Edit) 5 637.
[14] Liang J, Wang F, Dang C and Qian Y 2012 Inter J. Approx. Reason. 6 912.
[15] Zadeh L A 1965 Inform. Contr. 8 338.
[16] Shi L, Chen L and Li K 2012 J. Tianjin Norm Univ. (Natur. Sci. Edit.) 2 59.
[17] Wang T, Wang G and Chen Y 2012 Inform. Magaz. 6 47.
[18] Wang T 2012 Research on the Method of Network Public Opinion Analysis and Threat Estimation in Emergency (Zhengzhou: PLA Information Engineering University).
[19] Wang T, Wang G and Chen Y 2012 J. Inform. Engi. Unive. 8 482.
[20] Zhang Y 2015 Inform Sci. 11 100.
[21] Li W, Hua C and He W 2015 Inform Sci 9 93.
[22] Zhang F, Hua C and He W 2013 J. Chongqing Unive. Tech (Natur Sci) 12 123.
[23] Dai Y, Hao X and Guo Y 2010 Tech Res 5 60.
[24] Wang G and Li M 2017 Modern Inform. 1 41.
[25] Chen P and Hou T 2019 Inform. Sci. 5 115.
[26] Li C, Wang J and Lin C 2010 Appli Res. Comput. 9 312.
[27] Atanassovk K 1986 Fuzzy Sets. Sys. 1 87.
[28] Atanassovk K 1989 Fuzzy Sets. Sys. 1 37.
[29] Lin C, Li Y and Wang J 2011 J. Inform. Engi. Unive. 2 72.
[30] Wang J 2011 Research on Network Public Opinion Situation Analysis and Threat Estimation Technology Based on Intuitionistic Fuzzy Reasoning (Zhengzhou: PLA Information Engineering University).
[31] Zhang Y, Li H and Peng L 2017 J. Inform. 6 111.
[32] Xiao H, Chen H and Chen F 2019 Inform. Explor. 7 29.
[33] Chen T and Lin J 2013 Appli. Math. & Inform. Sci. 6 2335.
[34] Zhang C 2014 Research on Fuzzy Cellular Automata Model of Internet Public Opinion Transmission (Lanzhou: Northwest Normal University).
[35] Dang X, Zhang C and Hao Z 2014 Comp. Engi. 4 209.
[36] Dai J and Hang J 2012 J. Inform. 7 16.
[37] Yu W and Li J 2016 Inform. Explor. 4 7.
[38] Wang J, Li Yi and Huang J 2011 J. Inform. Engi. Univ. 10 585.
[39] Liu Y 2012 *Statist. Deci.* **2** 12.
[40] Zhang Q, Xie B and Zhang X 2015 *Comp. Sci.* **6** 70.
[41] Wang X 2015 *Analysis of User Attributes and Event Sentiment Based on Internet Public Opinion Data* (Shanghai: Shanghai University).
[42] Liu J, Bi Q and Li R 2016 *Inform. Theo. Prac.* **12** 31.
[43] Hu Y and Wang Y 2017 *Inform. Magaz.* **6** 111.
[44] Fan Z, Tian Y and Lv L 2017 *Inform. Sci.* **12** 28.