The University Network Public Opinion Monitoring System Based on Multi-Agent

Ji Weige[1, a] , Lv Qiang[2, b *]

[1]Hebei University of Science and Technology, China

[2]School of Economics and Management, Hebei University of Science and Technology, China

[a]jwg@hebust.edu.cn, [b]595215839@qq.com

Keywords: Multi-Agent; Network public opinion; Monitoring system.

Abstract. The article is base on the Multi-Agent theory, analyzing the structure of university network public opinion monitoring system, building the university network public opinion monitoring system based on Multi-Agent and introducing the effect of each Agent in detail. Putting forward a feasible scheme for university Network public opinion monitor.

Introduction

Internet created a new social space, and changing every aspect of social life quickly. Facing such a new society, The students express ideas and participate in the construction and management of schools through the Internet. In colleges and universities, teachers and students can pay attention to hot social issues constantly through the Internet. Then they can publish their own opinions, exchange ideas and increase participation in the decision-making process[1]. By this way, the network consensus form rapidly and it put forward a huge challenge to the management of colleges and universities.

The article uses the Multi-Agent technology in the university network public opinion monitoring system, and builds the structure of the system[2]. On this basis, the function of each Agent is analyzed in detail. We hopes to be able to promote the construction of the harmonious environment in university, and provides a feasible research method to network public opinion monitoring for colleges and universities[3].

The Structure of University Network Public Opinion Monitoring System

The university network public opinion monitoring system based on Multi-Agent uses a typical distributed network overall structure[4]. The Internet Agent Set analysis of processing large website information is called Concentrated Agent Set. Every University Agent Set is mainly responsible for school-related information network public opinion analysis and processing, and it's known as Dispersed Agent Set. The Internet Agent Set can release the latest information on network public opinion to University Agent Sets through the communication network and information platform[5]. And University Agent Sets can also share information and communicate with each other via the communication network and information platform.As shown in Fig. 1:

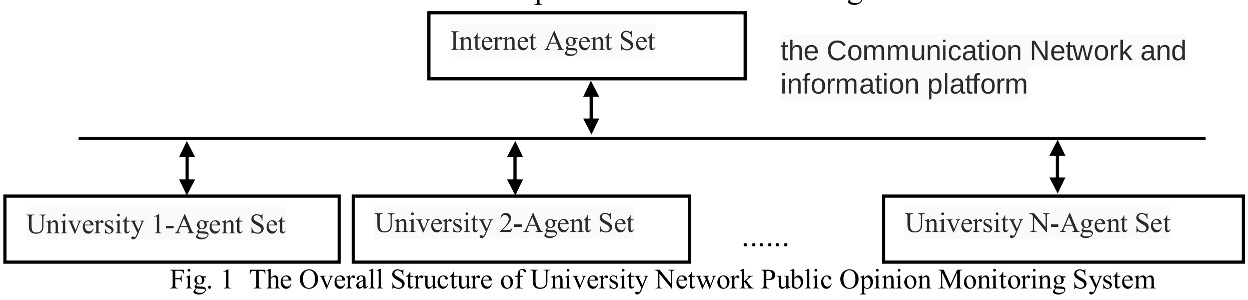

Fig. 1 The Overall Structure of University Network Public Opinion Monitoring System

In Fig. 2, the overall structure of the university network public opinion monitoring system based on Multi-Agent were refined. It describes the framework of the university network public opinion monitoring system based on Multi-Agent, and the Concentrated Agent Set is made up by Collection-Agent, Data-Agent and Analysis-Agent.

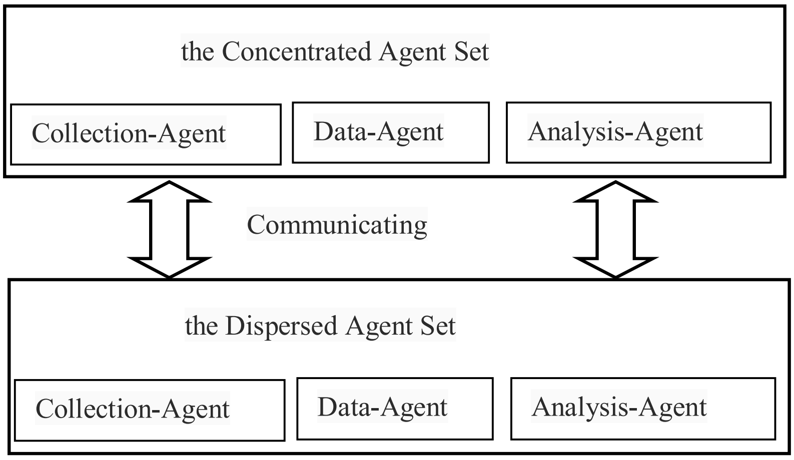

Fig. 2 The Framework of the University Network Public Opinion Monitoring System

The Dispersed Agent Set is responsible for public opinion network management within the school. It's also made up by Collection-Agent, Data-Agent and Analysis-Agent. Each Agent of the Dispersed Agent Set has the same functions of the Concentrated Agent Set, but the Concentrated Agent Set is used for the Internet and the Dispersed Agent Set is only for schools.

The Function of Each Agent on the University Network Public Opinion Monitoring System

The Function of Collection-Agent. The Collection-Agent use PageRank Algorithm to evaluate the importance of web pages[6], and then collecting the text of each web page. PageRank algorithm equation such as Eq. 1:

$$PageRank(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PageRank(p_j)}{L(p_j)}. \tag{1}$$

In the formula, p_i is the web page, PageRank(p_i) is the PageRank value of each web page. M(p_i) is the number of web pages to link. L(p_i) is the number of web pages to be linked. And d is the damping factor. 1-d is the probability of jumping to a new page. N is the number of all web pages.

A page's PageRank value calculated by the other pages' PageRank values. It's required that the original PageRank value of each page is 1. Collection-Agent calculate PageRank values of each page until all values tend to be stable. Then Fetching web pages that PageRank values are greater than the threshold value to extract text.

The Function of Data-Agent. Data-Agent sets the text which extracted by Collection-Agent into words, and using the TF-IDF algorithm gives the words different weights[7]. By this way, we turn the text into a space vector model[8].

The text Di turns into words such as Eq. 2:

$$W_i = \{w_{i1}, w_{i2}, w_{i3} ... w_{ik}\}. \tag{2}$$

TF-IDF formula such as Eq. 3:

$$v_{ik} = \frac{\log(tf_{ik}+1) * \log(idf_{ik})}{\sqrt{\sum_{i=1}^{N}[\log(tf_{ik}+1) * \log(idf_{ik})]^2}}. \tag{3}$$

In the formula, N is the number of all texts. tf_{ik} is the ratio of the number of the word w_{ik} occurrences of text D_i to the total number of all words in text D_i. idf_{ik} is the ratio of the number N of all texts to the number of texts that contain the word w_{ik}.

Calculating the weight for each word such as Eq. 4:

$$V_i = \{v_{i1}, v_{i2}, v_{i3} ... v_{ik}\}. \tag{4}$$

So the space vector model for text D_i such as Eq. 5:

$$D_i = \{(w_{i1}, v_{i1}), (w_{i2}, v_{i2}), (w_{i3}, v_{i3}) ... (w_{ik}, v_{ik}),\}. \tag{5}$$

The Function of Analysis-Agent. Firstly Analysis-Agent uses the SinglePass algorithm to classify space vector models by topics, and then uses cosine similarity to calculate the similarity between two space vector models such as Eq. 6[9]:

$$sim(D,T)=\frac{\sum_{i=1}^{n}d_i t_i}{\sqrt{(\sum_{i=1}^{n}d_i^{\ 2})(\sum_{i=1}^{n}t_i^{\ 2})}}\ . \tag{6}$$

In the formula, D and T are two space vector models that need to compare. d_i and t_i are weights of each word of two space vector models.

Secondly calculating the heat of every topic such as Eq. 7[10]:

$$h_{ij}=\frac{n_{ij}*10}{N_{ij}}*w_1+w_2*\log_{(r_{ij}+c_{ij})}(r_{ij}+w_3*c_{ij})\ . \tag{7}$$

In the formula, h_{ij} is the heat of topic i on day j. N_{ij} is the total number of reports on day j, and n_{ij} is the number of reports about the topic. Then r_{ij} is views of the topic, c_{ij} is comments of the topic. The symbols ω_1, ω_2, ω_3 are weighting factors.

At last, using Eq. 8 to choose heat rising topics.

$$H_i=\frac{h_{ij}}{\sum_{k=1}^{j-1}h_{ik}}*(j-1)\ . \tag{8}$$

In the formula, h_{ij} is the heat of the topic i on day j, h_{ik} is the heat of each day of j-1 days before day j. Then setting the threshold K and comparing with H_i and K, and choosing topics that H_i is greater than T.

The Application Example

Firstly, we used Collection-Agent to grab 20 web pages that published in Sina Weibo on June 28, 2015 randomly. And evaluating the importance of web pages: PR_1=1.053, PR_2=0.592, PR_3=0.875, PR_4=1.140, PR_5=0.920, PR_6=1.059, PR_7=0.973, PR_8=0.579, PR_9=0.547, PR_{10}=1.028, PR_{11}=1.400, PR_{12}=0.825, PR_{13}=0.663, PR_{14}=1.015, PR_{15}=0.920, PR_{16}=0.659, PR_{17}=0.939, PR_{18}=0.848, PR_{19}=0.419, PR_{20}=0.947. Then Fetching web pages that PageRank values are greater than 1 to extract text. After this we used Data-Agent to get space vector models such as Table 1:

Table 1 the Space Vector Models of Web Pages

The Space Vector Models
D_1={(laws,0.522),(society,0.476),(morality,0.355),(Anti-corruption,0.294),(responsibility,0.155),(supervision,0.140),(serious,0.053),(consciousness,0.051),(leadership,0.047),(discipline,0.047),(reform,0.045),(education,0.045)...}
D_2={(Peking,0.502),(Tsinghua,0.502),(universities,0.298),(examination,0.184),(students,0.161),(competition,0.138),(quality,0.115),(education,0.115),(weibo,0.069),(tradition,0.069),(society,0.069),(security,0.069),(champion,0.069)...}
D_3={(champion,0.461),(universities,0.441),(education,0.362),(entrance,0.362),(students,0.242),(Peking,0.202),(Tsinghua,0.202),(scholarship,0.121),(graduation,0.101),(financial,0.101),(emphasis,0.098),(competition,0.061)...}
D_4={(market,0.625),(stock,0.430),(drop,0.315),(rose,0.200),(regulatory,0.167),(policy,0.150),(securities,0.150),(normality,0.117),(transaction,0.116),(register,0.084),(reform,0.084),(capital,0.084),(society,0.084),(bubble,0.084)...}
D_5={(examination,0.417),(admissions,0.417),(champion,0.417),(universities,0.329),(evaluation,0.329),(independent,0.240),(students,0.240),(capacity,0.210),(score,0.165),(volunteer,0.120),(workplace,0.091)(Peking,0.074)...}
D_6={(candidate,0.380),(answer,0.324),(examination,0.324),(education,0.324),(rumors,0.286),(evaluation,0.286),(students,0.229),(review,0.172),(admission,0.115),(score,0.115),(work,0.092),(universities,0.092),(specification,0.065)...}

Secondly, we used Analysis-Agent to calculate the similarity between two space vector models such as:

$sim(D_1,D_2)$=0.069, $sim(D_3,D_1)$=0.042, $sim(D_3,D_2)$=0.579, $sim(D_4,D_1)$=0.035, $sim(D_4,D_2)$=0.006, $sim(D_4,D_3)$=0.005, $sim(D_5,D_1)$=0.037, $sim(D_5,D_2)$=0.336, $sim(D_6,D_1)$=0.013, $sim(D_6,D_2)$=0.376.

When the similarity is more than 0.3, we think the two models for the same topic. So that we could get three topics, such as: D_1 belongs to the first topic named "Anti-corruption", D_2, D_3, D_5, D_6 belong to the second topic named "The university entrance exam" and D_4 belongs to the third topic named "Stock".

After this,we fetched 100 web pages as the experimental data everyday On June 28 to July 2.And calculating the heat of three topics on each day such as:
h_{11}=0.427,h_{12}=0.394,h_{13}=0.468,h_{14}=0.486,h_{15}=0.424,h_{16}=0.467,h_{21}=0.548,h_{22}=0.568,h_{23}=0.508,h_{24}=0.605,h_{25}=0.527,h_{26}=0.646,h_{31}=0.486,h_{32}=0.466,h_{33}=0.373,h_{34}=0.407,h_{35}=0.450,h_{36}=0.388.

At last, we calculated the H of each topic, such as: H_1=1.062, H_2=1.172, H_3=0.889. We set the threshold value T for 1. If H is greater than 1, it's said that the heat of the topic will increase. So the topic "Anti-corruption" and "The university entrance exam" is likely to form network public opinion.

Summary

The article uses the Multi-Agent technology in the university network public opinion monitoring system, provides to the university network public opinion management strategy, which is beneficial to improve the level of university network public opinion emergency management, maintain campus security and stability. The system can analyze and discover potential of university network public opinion, and provide early warming to managers of universities, set up the consciousness of crisis, improve the ability of public opinion management and the management system.

Acknowledgements

This paper is supported by the project from national natural Foundation of China (71301044). It also gets help from the project of Hebei Education Department of China (201418, ZD2014027).

References

[1] Zhang Jiyan, Exploration of College Students' Network Behavior of University under the Network Public Opinion, Higher Education Forum, 1 (2015) 12-14.

[2] Nicholas R.Jennings, On Agent-Oriented Software Engineering, Artificial Intelligence, 117 (2000) 277-296.

[3] Zen Runxi, Research on Information Resource Sharing of Network Opinion, Journal of Intelligence, 28 (2009) 187-191.

[4] Zhang Qi, Distributed Collaborative Design on Network, Xi'an: Xi'an University of Science and Technology, 2005.

[5] Zhou kai, Yan Yan, Song Bin, A Web-based Information Exchanging and Public Participation Framework Under Information, UPI, 27 (2012) 103-107.

[6] Huang Decai,Qi Chunhua, PageRank Algorithm Research, Computer Engineering, 32 (2006) 145-162.

[7] Huang Chenghui, Yin Jian, Hou Fang, A Text Similarity Measurement Combining Word Semantic Information with TF-IDF Method, Chinese Journal of Computers, 2011,34(5), 856-864.

[8] Xiu Chi, The Research and Implementation of Method for Domain Chinese Word Segmentation, Beijing: Beijing University of Technology, 2013.

[9] Zhou Gang, Zou Hongcheng, Xiong Xiaobing, Huang Yongzhong, MB-SinglePass: Microblog Topic Detection Based on Combined Similarity, Computer Science, 39 (2012), 198-202.

[10] Yin Fengjing, A Research of Internet Public Opinion Monitoring-oriented Hot Topic Detection Technology, Changsha: Graduate School of National University of Defense Technology, 2010.