Text Mining-based Analysis of Public Opinions to Environmental Public Services in Jiangsu Province

Cheng Yu[1] , Yang Li[1,2,*], Yu Qianqian[1],Wang Min[1]

[1] School of Economics, Nanjing University of Posts and Telecommunications, Nanjing 210046, China

[2] College of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210046, China

*Corresponding author: 1063385356@qq.com

Abstract: In the research project, Python language was used to write program and capture data from Sina Weibo; Besides, text mining was adopted for analysis of public opinions on environmental public services in Jiangsu Province. Research showed that the public tend to pay more attention to noise pollution, community environmental health and hazardous chemicals. Moreover, they were eager to know how serious environmental problems were and what should be to blame for such problems. High negative emotion indexes on community environmental health, environmental governance, environmental protection supervision and air quality were common among the public.

1.Introduction

Environmental public services refer to the institution where the government offers environmental public facilities in line with public demands [1]. As to satisfaction evaluation of environmental public services, questionnaire survey [2] and the establishment of satisfaction index system [3] on human settlement environment are common practices with foreign scholars. However, limited release volume and uncertain quality are often big challenges to questionnaire survey. Besides, human settlement environment evaluation often goes on a macro basis, making it difficult for specific problems to be discovered. Various social networking tools, represented mainly by Weibo, come as a new supplement to the evaluation of environmental public services. For one thing, it is possible for the public to see their intuitive feelings to local environmental public services spread through the social network; for another thing, massive information from social network communities comes as a real and immediate reflection of the public's degree of satisfaction to local environmental public services. This, in turn, will urge the local government to make immediate and effective rectification to problems in environmental public services.

The research project took Jiangsu Province as a typical region and adopted text mining method to carry out evaluation research on resident satisfaction on environmental public services by introducing new and empirical research methods related to environmental services.

2.Data collection and preprocessing

2.1 Data source and collection

The research project was conducted based on Sina Weibo popular among users and data involved in the project sourced mainly from user comments to environmental public services. Python language was used for writing web crawler program to capture Weibo data related to environmental quality and environmental public services. Keywords are shown in Table 1.

Table 1. keywords to environmental public service satisfaction evaluation

keywords	
environmental protection facilities	environmental protection propaganda
environmental pollution	environment and health
environmental protection	environmental health
hazardous chemicals	waste transportation
environment	water pollution
waste collection	drinking water
waste incineration	wastewater
ozone pollution	nuclear radiation
nuclear pollution	air quality
rubbish	waste sorting
PM2.5	haze
noise	soil

2.2 Text pro-preprocessing

The project began with original data cleaning, so as to exclude invalid and abnormal texts. Then "Jieba" package was used for word segmentation of valid data, sorting out of corpus on environmental evaluation and inputting user-defined dictionaries. This way, professional words will not be mistakenly segmented and segmented word texts will be established. Finally, symbols, figures and non-content auxiliary words in segmented word texts were screened, and final data files were obtained. Text pro-preprocessing went in the following procedure (see Figure1). After the data cleaning, totally 14,687 pieces of Weibo text data were collected.

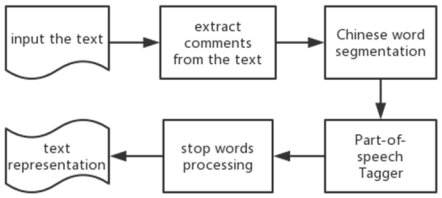

Fig 1. Text pro-preprocessing

3.Text mining and analysis of environmental services

Analysis to texts related to environmental services mainly comprises three parts: the first one is about overall analysis to public opinions to environmental services in Jiangsu based on TFIDF （term frequency–inverse document frequency）[4]; the second one is about analysis, based on LDA topic model, of the public's focus of attention to environmental services and topic classification; the third one is about emotion analysis, based on "SnowNLP", of various environmental service topics, in an effort to dig into Jiangsu residents' emotion tendency to various environmental public services in local communities.

3.1 Overview of Jiangsu environmental services

By studying, selecting, calculating and sequencing TFIDF values, hot words associated with environmental public services in Jiangsu were collected. TFIDF calculation follows the formula below:

$$TFIDF = TF \times IDF = \frac{N(word:text_{word})}{|text_{word}|} \times log\frac{|corpus|}{|\{d|word \in d\}|} \quad (1)$$

After calculation of the TFIDF value of each word, keywords with relatively higher TFIDF value were listed as follows (see Table 2).

Table 2. TFIDF ranking on environmental quality evaluation

words	**ranking**	**words**	**ranking**
leaking stoppage	1	bid inviting	16
power supply	2	prevention	17
soundboard	3	monitoring station	18
electric generator	4	wetland	19
hydrargyrum	5	fresh-caring	20
particles	6	circulation	21
recycling	7	testing	22
humidifier	8	legislation	23
generator set	9	bribery	24
dehydrator	10	gasoline	25
remove	11	illness	26
solidify	12	cold	27
emulsion	13	noise	28
greenhouse	14	ventilation	29
polyester	15	paint	30

Public opinions to environmental public services in Jiangsu Province fall into the following categories: (1) pollution arising from industrial production and decoration, such as leaking stoppage, power supply, electric generator, greenhouse and so on; (2) air pollution, such as particles, humidifier, gasoline, illness, cold, noisy ventilation, paint and so on; (3) environmental monitoring and remediation, such as monitoring station, testing, legislation, bribery and so on; (4) noise pollution, such as soundboard, electric generator, noise and so on.

According to the foregoing analysis: (1) Jiangsu is characterized by well-developed industry which is mainly to blame for air pollution, noise pollution, light pollution and other factors influential to local environmental quality. （2）Residents in Jiangsu Province are concerned more about noise and air pollution. (3) Where the environmental quality problem arises, people will tend to blame it on weak supervision and often resort to remediation and legislation.

3.2 LDA topics of Jiangsu environmental services

With the adoption of LDA（Latent Dirichlet allocation）topic model, the research project categorized the environmental public services in Jiangsu. Generation of the model follows this process: a certain topic z is given and the corresponding document is generated based on the given topic; all words involved in the document originate from the same topic; if the number of topics totals to k, those topics will be expressed in k, z1, z2, z3… zk respectively; the generation probability for each topic will be

expressed in p (z_1), p(z_2), p (z_3),⋯p(z_k); $p(w_i|z_k)$ stands for the probability that the word w_i will appear in the documents related to topics that count k. The probability to generate document d will be:

$$p(d) = p(z_1)\prod_{i=1}^{N} p(w_i|z_1) + p(z_2)\prod_{i=1}^{N} p(w_i|z_2) + \cdots + p(z_k)\prod_{i=1}^{N} p(w_i|z_k) \quad (2)$$

After non-supervision topic model learning to all the captured word-segmented documents, topics and keywords related to public opinions on Jiangsu environmental public services were collected, shown in Table 3:

Table 3. Topics and keywords related to public opinions on the environment

Topic No.	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
term1	noise	corruption	chemicals	pipeline	wetland	factory	recycle	ozone
term 2	rumble	worker	transfer	cough	inshore	mask	detain	straw
term 3	diesel oil	plastic	permit	washing	oil field	cancer	identify	offgas
term 4	assessment	graceful	explosion	shock	reservoir	academic	suffer	purify
term 5	machine set	street	governance	reservoir	riverway	professor	supervise	particle
term 6	decontaminat ion	waste soil	fire control	cancerogen	sea areas	complaints	director	breathe
term 7	electric generator	high temperature	inspection	win bidding	Lake Taihu	governance	telephone	volatility
term 8	resident disturbance	garbage truck	warehousing	drinking water	water control	water resource	leak stoppage	visibility
term 9	construction sites	sanitation work	gasoline station	water purification	sewage charge	responsibilit y linkage	rule by law	humidifier
term 10	electric generator	environment al sanitation	hazardous chemicals	Treatment plant	blue-green algae	purified water	rule by law	heavy fog

According to LDA topic model: (1) Existing environmental problems in Jiangsu include living water pollution, water pollution, air pollution and urban road sanitation; (2) Not only will netizens vent out negative emotions through social network, but also dig into causes for problems; (3) The public are concerned about neighboring facilities like chemical plants, chemical industry park and so on. This is because they are poor in knowledge of hazardous chemicals and have no confidence in controlling measures by enterprises concerned and local governments.

3.3 Emotion analysis of Jiangsu environmental services

Text emotion refers to the process in which subjective and emotion texts are analyzed, processed, generalized and reasoned. Firstly emotional thesaurus from HowNet is adopted for corpus practice; then SnowNLP, the third-party library of Python, is used for emotional tendency analysis; finally, the negative emotional probability for each topic from various Weibo posts is weighed, thus negative

emotion indexes corresponding to various topics are finalized (see Figure 2).

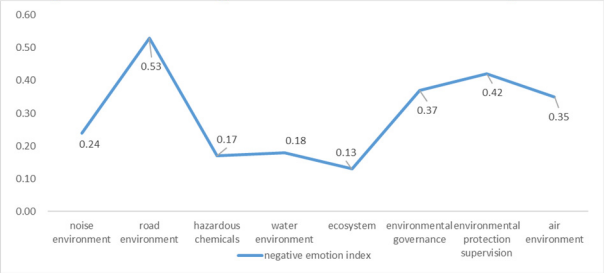

Fig 2. Line chart for Weibo public opinion emotion index

As indicated by the emotion analysis, topics with relatively high negative emotion index include road environment, environmental governance, environmental protection supervision and air environment. Noise pollution features an average negative index; hazardous chemicals, water environment and ecological water system are characterized by a relatively low negative index.

As road environment and air quality are closely linked to people's daily life, users are relatively more sensitive. Therefore, the public are more likely to get angry at poor supervision by environmental protection authorities and governments. Once being added fuel to the fire by media, pollution accidents caused by weak supervision will trigger public anger, which will finally erupt and spread through social networks. Noise pollution has a limited effect on public life; besides, hazardous chemicals, water environment, ecosystem and other relevant accidents often happen mostly in suburbs or remote areas. Therefore, users are not sensitive to such accidents, with a small negative effect.

4.Conclusions

(1) Jiangsu residents are concerned more about the following environmental public services: noise environment, community environmental health, hazardous chemicals, living water environment, ecological water environment, environmental governance, environmental protection supervision and air quality. Of them, they pay the greatest attention to water pollution, air pollution, community environmental health and other factors closely associated with their daily life.

(2) Topics of public opinions on environmental public services have its own characteristics, where the public show interest in not only environmental problems but the nature and reasons for such problems.

(3) Once a certain environmental problem in social network communities draws more attention from the public, it will see a relatively high negative emotion index. The public seldom voice their positive emotion opinions to public service through social networks; unable to find the right listeners to vent out their emotion, however, the public tend to voice their negative opinions to environmental public services through social networks.

Acknowledgments

This paper is jointly funded by National Social Science Foundation(16BRK024), Social Science Foundation of Jiangsu Province (15GLB016), Jiangsu University Major Program of Philosophy and Social Science (2017ZDIXM123), and 1311 Talent Program of Nanjing University of Posts and Telecommunications.

References

[1] Zheng Yan, Public Opinions and Publicness – Citizen Discourse in Weibo and Reflection[J]. *Literature and Art Studies*, 2012(04):26-34.

[2] Zi Fei, Yang Zhihui, Zhang Fan, Tian Hao, Wang Guangxin, Wu Jianping, Liu Yang, Fang Gang, Xiang Jinjing, Fang Ran. Investigation Reports on Degree of Environmental and Life Satisfaction in 10 Chinese Cities[J]. *Journal of Beijing Forestry University (social science edition)*, 2012,11(04):1-7.

[3] Liu Xingyun, Jiao Dongdong, Zhang Xiongkuan, Zhu Tingshao. About Evaluations to Government Environmental Protection Projects based on Public Satisfaction – Taking Beijing for Example[J]. *Journal of Arid Land Resources and Environment,* 2018, 32(01):50-55.
[4] Liu L, Peng T. Clustering-based method for positive and unlabeled text categorization enhanced by improved TFIDF. *Journal of Information Science and Engineering* 2014; 30(5):1463–1481.