Air quality perception satisfaction and influence factors analysis in Shandong, China

Yan Zheng, Yong Sun*, Min Ji and Xin Zhang
School of Geodesy and Geomatics, Shandong University of Science and Technology, China

*Corresponding author: yong_sun2020@sdust.edu.cn

Abstract. With the development of the economy and the improvement of the quality of life, people's sensitivity and the willingness to the air quality have also increased. In order to analyze the public’s perceived satisfaction with air quality, this paper processed and cleaned up the data of more than 20,000 air pollution complaints in Shandong Province, and calculated the air quality perceived satisfaction by constructing an emotional words library in the field of air quality. Based on the calculation results of air quality perception satisfaction in Shandong Province from 2017 to 2018, and then combined with IoT monitoring data, industrial development level and GDP data, the factors affected air quality perception satisfaction were analyzed. The results show that: (1) The public’s perceived satisfaction with air quality is not directly related to the concentration of objective pollutants in the air. (2) In areas where industrial enterprises are densely distributed, the perceived satisfaction with air quality tends to be worse. (3) In cities with high levels of economic development, the public’s perceived satisfaction with air quality tends to be worse.

Keywords: air quality, emotion analysis, satisfaction, Shandong province.

1. Introduction
Perceived satisfaction of air quality is the reflection of objective air quality through subjective direct experience [1]. The public’s perceived satisfaction with air quality reflects the distribution of air pollution and reflects people’s satisfaction with the surrounding air quality. Therefore, in order to better assist relevant government departments in environmental governance, it is of great significance to analyze the satisfaction of air quality perception and its influencing factors.

Initially, perceived quality was obtained through questionnaire surveys to obtain the public’s evaluation of products. Olsen first proposed that perceived quality was consumers' perception of product quality [2]. Based on consumers' perception of product quality, air quality perception has begun to be studied by many scholars. Zhang took Haidian District as the research area and found that residents can perceive pollution, and their willingness to perceive has been increasing in recent years [3]; Zhang found that social structure status will affected people’s air pollution exposure risk [4]; Feo found that the smell of garbage and the location change of garbage treatment plants had a direct impact on residents’ perception of air quality [5].
With the development and popularization of network technology, the public is more inclined to express opinions on the internet. In this environment, some platforms for complaints and suggestions about air quality have emerged. There are tens of thousands of comment data on these platforms, which are dynamically updated in real time. The comments contain various emotional colors and emotional tendencies of the public [6]. Using comment data for opinion mining, that is, sentiment analysis [7,8], can obtain the public’s perception of atmospheric quality, and then provide a large amount of comprehensive first-hand data for the research of air quality perception satisfaction. At present, the research on sentiment analysis is very extensive. Commonly used methods include dictionary-based sentiment analysis and machine learning-based sentiment analysis. Sentiment analysis based on machine learning is suitable for dividing emotions into several categories, the processing process is very complicated and the accuracy rate is low [9,10]; dictionary-based sentiment analysis can more accurately calculate the sentiment intensity of the text by adding, deleting, and supplementing the existing sentiment dictionary, with higher calculation accuracy and efficiency [11,12]. The dictionary-based sentiment analysis method can also calculate the sentiment intensity by constructing a domain dictionary. Lin conducted sentiment analysis on hotel review data by constructing a domain dictionary for hotel reviews, compared with basic sentiment dictionary, it had better domain expression and usability, which could improve the calculation accuracy of sentiment analysis [13]; Xi constructed a domain sentiment dictionary for product reviews, and the results showed that it had a good effect on the actual product review data set [14]; Zheng through the construction of an emotional vocabulary in the field of air quality, the satisfaction of air quality perception had been studied, and good results had been achieved [15], but did not analyze the influencing factors of air quality perception satisfaction.

Through the above research, it can be seen that the air quality perception satisfaction and its influencing factors play an important role in air quality evaluation. The traditional paper questionnaire survey method for air quality perception satisfaction research is more costly and has a narrow coverage, not conducive to dynamic tracking description of air quality perception satisfaction and its influencing factors. In this paper, the sentiment analysis method based on the domain sentiment vocabulary is used to calculate the public's perceived satisfaction with air quality, and then the calculated air quality perceived satisfaction is combined with IoT monitoring data, industrial enterprise development and GDP data to calculate the air quality perceived satisfaction. Thereby real-time dynamic analysis of air quality perception satisfaction and its influencing factors is realized.

2. Method

2.1. Data collection

The research data in this paper is the public complaint page of Shandong Environmental Public Prosecution (letters and visits) platform crawled by the crawler module, and 22,269 complaint data from 2011 to 2018 were captured through XPath technology, including the following fields: number, classification, handling status, date, page views, title, content, response status. Because the research content only need to analyze the public's perception and satisfaction of the surrounding air quality in the complaint data, the classification, complaint time, title and content fields were extracted from the data set. The data format is shown in Table 1.
Table 1. Complaint data.

| Classification | Date       | Title          | Content                                                                 |
|----------------|------------|----------------|-------------------------------------------------------------------------|
| Air Pollution  | 20180911   | Stealing sewage| The shale brick factory secretly discharged sewage again. The Environmental Protection Agency does not care! |
|                |            |                | The ditch behind the west end of the village was stinking all day long, and nearby villagers complained. Please the relevant departments to take care of it. |
| Water pollution| 20171115   | The ditch stinks| The community smells of ash and sulfur dioxide every day, so I dare not open windows for ventilation and go out. I hope to lead the field investigation and not listen to the factory and monitoring data. |
| Air Pollution  | 20180311   | Waste gas pollution| Shengma Company sneaks a smell of chemical burnt and pesticide every day and night. No one asks at night when harmful gases are emitted every day. |
| Air Pollution  | 20181003   | Odor pollution |                                                                          |

2.2. Research framework of air quality perception satisfaction and influence factors

Based on the basic emotional vocabulary, this paper constructed an emotional dictionary in the field of air quality, which together constituted an emotional vocabulary in the field of air quality, and used the method based on the emotional words library in the field of air quality to calculate the air quality perception satisfaction, so as to analysis the influence factors of air perception quality satisfaction. The research framework is shown in Figure 1.

Figure 1. Research framework of air quality perception satisfaction and influencing factors.

2.2.1. Basic emotional words library. Basic emotional words library consists of basic emotional lexicon and modifier dictionary. The basic emotional lexicon is a Chinese sentiment vocabulary resource compiled and annotated by the Information Retrieval Laboratory of Dalian University of Technology. The parts of speech included are: verb, noun, adjective (adj), network words (Nw), adverb (adv), idiom. Each word is marked with the corresponding emotional polarity, "0" means neutral, "1" means commendatory, and "2" means derogatory. Specific examples are shown in Table 2.
Table 2. Example of basic emotional library.

| Word       | Part of speech | Classification | Intensity | Polarity |
|------------|----------------|----------------|-----------|----------|
| refreshing | adj            | Praise         | 5         | 1        |
| fierce     | adj            | Fear           | 3         | 0        |
| sunny      | adj            | Praise         | 5         | 1        |
| pandemonium| idiom          | Derogatory     | 3         | 2        |

The modifier dictionary includes a negative word dictionary and a degree adverb dictionary. Negative words include: "Do not", "No", "Is not", etc.; adverbs of degree (Table 3) are divided into five levels: "owe", "slightly", "relatively", "very" and "extremely". This paper assigned values to them according to experience: "owe" and "slightly" had a certain weakening effect, which were assigned values of 0.6 and 0.8 according to the degree of weakening. "relatively", "very" and "extremely" had a certain emphasizing effect on emotional words, which were assigned values of 1.2, 2 and 3 respectively.

Table 3. Degree adverb dictionary.

| Number | Level    | Word                              | Weights |
|--------|----------|-----------------------------------|---------|
| 1      | extremely| special, incredible               | 3       |
| 2      | very     | extraordinarily, good             | 2       |
| 3      | relatively| more, relatively                  | 1.2     |
| 4      | slightly | slightly, a little                | 0.8     |
| 5      | owe      | not a little bit, mild             | 0.6     |

2.2.2. Emotional words library in the field of air quality. The emotional words library in the field of air quality includes a basic emotional words library and an emotional dictionary in the field of air quality. For the field of air quality, it is impossible to accurately calculate the public's perception tendency by relying only on the basic emotional words library. Therefore, this paper constructed an emotional words library about air quality through Chinese word segmentation, thematic words extraction, and emotional intensity labeling.

(1) Chinese word segmentation
   In this paper, the Chinese word segmentation method based on the jieba segmentation added to the user-defined dictionary was used to segment the complaint data.
   ① Based on the training of the People's Daily corpus and other resources, the dictionary generated a prefix tree, and a directed acyclic graph (DAG) of all possible words was generated based on the prefix dictionary.
   ② Dynamic programming found the path with the greatest probability and found the largest segmentation combination based on word frequency.
   ③ For unregistered words, a Hidden Markov model (HMM) based on the ability of Chinese characters to form words was used, and the Viterbi algorithm was used for calculation.
   ④ Add a custom dictionary (for example: if the word segmentation phrases are "pollution" and "serious", and the custom word "serious pollution" is added, then the word "pollution" and "serious" will not be separated during word segmentation).

(2) Thematic words extraction
   According to the part of speech in the emotional vocabulary body of Dalian University of Technology, traverse the word segmentation phrase, extracted the words that meet the emotional part of speech, and used manual extraction to extract thematic words in the air quality field.
(3) Emotional intensity labeling

According to the emotional intensity labeling criteria of emotional words in the basic sentiment
dictionary, the emotional intensity of air quality topic words was marked by means of manual
annotation by experts. The domain dictionary constructed through the above steps includes four
indicators of word, part of speech, intensity and polarity. The results are shown in Table 4.

**Table 4. Example of emotional words library in the field of air quality**

| Word          | Part of speech | Intensity | Polarity |
|---------------|----------------|-----------|----------|
| pungent       | adj            | 5         | 2        |
| clean         | noun           | 3         | 1        |
| can't breathe | idiom          | 7         | 2        |
| scenery       | idiom          | 7         | 1        |

2.2.3. *Air quality perception satisfaction*. In this paper, dictionary-based sentiment analysis was used
to calculate the air quality perception satisfaction of the complaint texts from 2017 to 2018. First,
segmented the text; then used formula (1) to calculate the air quality perception satisfaction of the
clause; finally, taken the minimum value of the clause air quality perception satisfaction as the final air
quality perception satisfaction of the sentence.

\[
E(P) = E(PW) \times E(NA) \times E(PA)
\]  

(1)

3. Results

3.1. Result analysis

Using the above sentiment analysis method, the basic sentiment dictionary and the expanded sentiment
dictionary were used to calculate 5,000 artificially-labeled sentiment value data in the complaint data.
The data with the sentiment value calculation error within the range is regarded as correct. Data in the
range is considered as an error. After experimental calculation, the accuracy is as high as 92.2%,
which can meet the experimental requirements. Some experimental results are shown in Table 5.

**Table 5. Partial experimental results.**

| Number | Text                                                                 | Polarity |
|--------|----------------------------------------------------------------------|----------|
| 1      | Factory sewage caused death.                                         | -9       |
| 2      | Wanton discharge of toxic gases, polluting the environment.          | -7       |
| 3      | Untreated exhaust gas.                                              | -1       |
| 4      | The smell makes me sick and vomiting.                               | -7       |
| 5      | Some companies spray paint overnight, seriously polluting the surrounding atmospheric environment. | -5       |

In this paper, using the complaint data of Shandong Province from 2017 to 2018, using the
emotional analysis method based on dictionary, calculated the air quality perception satisfaction, and
located according to the location information in the data, and made statistics on prefecture level cities,
calculated the average value of annual air quality perception satisfaction, and drew the distribution
map of air quality perception satisfaction in Shandong Province, as shown in Figure 2. The smaller the
numerical value in the figure, the worse the public's satisfaction with air quality.
It can be seen from the figure that Weifang, Yantai, Linyi, and Qingdao have poor air quality perception satisfaction, while Weihai and Rizhao two coastal cities have better air quality perception satisfaction. According to the statistical results of the main business income of industries in 17 cities in Shandong Province by the Yantai Statistics Bureau/Weihai Statistics Bureau, it is found that Weifang, Yantai, Linyi, and Qingdao are among the top five in Shandong Province by industrial income, and industrial enterprises are relatively developed, and Weihai and Rizhao are the last five in Shandong’s industrial income. The operation of industrial enterprises needs to burn a large amount of coal. The sulfur dioxide produced by coal combustion will seriously pollute the atmospheric environment, making residents’ satisfaction with air quality tend to be worse. Through further analysis of the complaint data, it is found that most of the complaints are about the dust and pungent gas emitted by industrial enterprises such as chemical plants and machinery companies. Therefore, the more serious the air pollution is perceived by the public in places where factories are densely populated. This is in line with the actual situation. Environmental protection departments can also take targeted measures to prevent and control air pollution, implement various environmental protection policies, and promote sustainable economic development.

3.2. Analysis the influence factors of air quality perception satisfaction
In order to further study the relationship between air quality perception satisfaction and objective pollutants in the atmosphere, this paper extracted the top 18 pollutants in the complaint data, including dust, waste gas, sewage, black smoke, coal, flue gas, soot, dust, smoke, paint, rubber, oily smoke, heavy smoke, feces, plastic particles, formaldehyde, smog, waste residue. For the five types of pollutants: paint, rubber, oil fume, feces, and formaldehyde, the public can perceive the foul smell they produce, but it has nothing to do with the pollutants in the ambient air quality standards. This kind of complaint data is eliminated to reduce the impact of irrelevant data on the results. Then this paper used correlation analysis to calculate air quality perception satisfaction with AQI, PM$_{2.5}$, PM$_{10}$, SO$_2$, CO, NO$_2$, and O$_3$ in the air. The experimental data analysis results were shown in Table 3.

| Objective pollutants | Variable | AQI | PM$_{2.5}$ | PM$_{10}$ | SO$_2$ | CO | NO$_2$ | O$_3$ |
|----------------------|----------|-----|------------|-----------|--------|----|-------|------|
| Air quality perception satisfaction Correlation coefficient | 0.074 | 0.002 | 0.052 | 0.111 | 0.053 | 0.067 | 0.237 |
| Significance | 0.778 | 0.994 | 0.884 | 0.671 | 0.841 | 0.800 | 0.359 |
From the results in Table 6, it can be seen that the air quality perception satisfaction is not directly related to the concentration of objective pollutants. In order to explain this phenomenon, this paper sorted the air quality perception satisfaction and AQI index of each city in order from low to high. The better the air quality perception satisfaction and the lower the AQI index, the closer the ranking is to 1, and then the ranking is poor. If the calculation result is positive, it means that the public's perception of air quality tends to be better. Otherwise, the perception of air quality satisfaction tends to be worse. Finally, the analysis was carried out based on the GDP data of each city in Shandong Province in 2018, as shown in the Figure 3, it was found that cities with higher GDP data tend to have worse air quality perception satisfaction, that was, the higher the public's living standards, the higher the requirements for the surrounding air quality.

![Figure 3. Perceptual difference and GDP correlation analysis.](image)

4. Conclusions
Through the calculation of air quality perception satisfaction and the analysis of influencing factors on the complaint text data, this paper drew the following conclusions: In the densely distributed areas of industrial enterprises, the air quality perception satisfaction tends to be worse; the concentration of objective pollutants in the air has no direct effect on the air quality perception satisfaction; cities with higher GDP, that is, cities with higher public income, have higher public requirements for air quality, which leads to a worse public perception of air quality. Based on the above conclusions, relevant government departments can control regional air quality in a targeted manner. While economic development, especially industrial development, they must pay more attention to the pollution emission control of industrial enterprises, so as to improve the public's living standards while increasing the overall public happiness.

Acknowledgments
This work was financially supported by the Major Scientific and Technology Innovation Projects of Shandong Province (2019JZZY020103).

References
[1] J. Li, Research on the Relationship among Perceived Air Quality, Public Satisfaction and Environmental Behavior Intention. China Jiliang University, 2016:1-96.
[2] J.C. Olson, Jacoby. Research of Perceiving Quality, Emerging Concepts in Marketing, 1972, (9):220 -226.
[3] J.Zhang, Y. Sun, D.L. Chen, Public understanding of smog pollution: A survey on the residents’ perception of air pollution in HaiDian District. Studies in Science of Science, 2017(04):14-22.

[4] W.X. Zhang, H.Z. Mu, H.M. Fan, Analysis on the difference of social status of air pollution exposure risk: a case study of Liaoning Province. Environmental Pollution & Control, 2017, 39(04):444-450.

[5] G.D. Feo, S.D. Gisi, I.D. Williams, Public perception of odour and environmental pollution attributed to MSW treatment and disposal facilities: A case study. Waste Management, 2013, 33(4):974-987.

[6] Y.Y. Zhao, B. Qin, T. Liu, Sentiment analysis. Journal of Software, 2010, 21(8):1834-1848.

[7] X. Li, C. Wu, F. Mai, The Effect of Online Reviews on Product Sales: A Joint Sentiment-Topic Analysis. Information & Management, 2018, 56(2):172-184.

[8] A. Trilla, F. Alias, Sentence-Based Sentiment Analysis for Expressive Text-to-Speech. IEEE Transactions on Audio Speech & Language Processing, 2013, 21(2):223-233.

[9] G.L. Chen, Microblog Sentiment Analysis Basing on Emotion Dictionary and Semantic Rule. Information Research, 2016(02):1-6.

[10] J.W. Sun, X.Q. Lv, L.H. Zhang, On sentiment analysis of chinese microblogging based on lexicon and machine learning. Computer Applications and Software, 2014, 31(7):177-181.

[11] C.G. Zhang, P.Y. Liu, Z.F. Zhu, A sentiment analysis method based on a polarity lexicon. Journal of Shandong University (Natural Science), 2012, 47(3):50-53.

[12] F. Yang, Y.D. Wu, X.Y. Wang, Sentiment Analysis of Chinese Hotel Reviews Based on the Extended basic Dictionary. Journal of Hubei University of Technology, 2019, 34(01):109-112.

[13] J.H. Lin, Y.M. Zhou, A.M. Yang, Building of domain sentiment lexicon based on word2vec. Journal of Shandong University (Engineering Science), 2018, 48(03):40-47.

[14] Y.H. Xi, Construction of Domain-specific Sentiment Lexicon in Product Reviews. Journal of Chinese Information Processing, 2016, 30(5):136-144.

[15] Y. Zheng, Y. Sun, M. Ji, Research on atmospheric quality perception satisfaction based on domain emotion lexicon. Environmental Pollution & Control, 2020, 42 (09): 1182-1184+1190.