applied sciences

Article

Can Urban Environmental Problems Be Accurately Identified? A Complaint Text Mining Method

Yaran Jiao [1,2], Chunming Li [1,*] and Yinglun Lin [3]

[1] Key Lab of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China; yrjiao@iue.ac.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] College of Resources and Environment, Fujian Agriculture and Forestry University, Fuzhou 350002, China; yllin@iue.ac.cn
* Correspondence: cmli@iue.ac.cn

Featured Application: This study establishes a framework for Chinese text mining of civil environmental complaints to provide a technical reference for the analysis of massive environmental complaint text data.

Abstract: With the popularization of social networks, the abundance of unstructured data regarding environmental complaints is rapidly increasing. This study established a text mining framework for Chinese civil environmental complaints and analyzed the characteristics of environmental complaints, including keywords, sentiment, and semantic networks, with two–year environmental complaints records in Guangzhou city, China. The results show that the keywords of environmental complaints can be effectively extracted, providing an accurate entry point for solving environmental problems; light pollution complaints are the most negative, and electromagnetic radiation complaints have the most fluctuating emotions, which may be due to the diversity of citizens' perceptions of pollution; the nodes of the semantic network reveal that citizens pay the most attention to pollution sources but the least attention to stakeholders; the edges of the semantic network shows that pollution sources and pollution receptors show the most concerning relationship, and the pollution receptors' relationships with pollution behaviors, sensory features, stakeholders, and individual health are also highlighted by citizens. Thus, environmental pollution management should not only strengthen the control of pollution sources but also pay attention to these characteristics. This study provides an efficient technical method for unstructured data analysis, which may be helpful for precise and smart environmental management.

Keywords: environmental complaint; text mining; semantic network; sentiment analysis; sustainable cities

Citation: Jiao, Y.; Li, C.; Lin, Y. Can Urban Environmental Problems Be Accurately Identified? A Complaint Text Mining Method. *Appl. Sci.* **2021**, *11*, 4087. https://doi.org/10.3390/app11094087

Academic Editor: Miltiadis D. Lytras

Received: 17 March 2021
Accepted: 28 April 2021
Published: 29 April 2021

1. Introduction

Environmental quality has become a critical factor for improving urban sustainability [1]. In the era of big data, smart cities provide citizens with a better living environment, which has become an emerging model of world city development. Its essence lies in the high integration of informatization and urbanization. With the rapid development of information technology and the increase in citizens' environmental awareness [2], it is more convenient to make environmental pollution complaints with the help of mobile phones and social networks. Citizens are more active in expressing their subjective feelings about environmental pollution. For example, in 2019, China's "12369" environmental protection reporting network management platform received more than 530,000 environmental complaints records from the public, of which Guangdong Province ranked second. Environmental complaint data are unstructured text data, which have different data analysis methods from traditional environmental sensor networks (such as the air quality monitoring network or water pollution monitoring network); furthermore, the density of environmental complaints is much higher than that of any of the current environmental sensor network sites. Massive environmental complaints have produced huge text data containing rich information, such as the characteristics of the pollution source, the information of stakeholders, and the perception regarding the complainants.

However, previous studies of environmental complaints mostly focused on correlating environmental complaints with socio–economic factors or individual features, including economic development, geographical location, household income, literacy rate, environmental management, age, gender, education quality, perception, which played significant roles in determining civil environmental complaints. For example, Dasgupta and Wheeler [3] evaluated the influencing factors of civil environmental complaints based on an econometric model, which proved basic education has a significant effect on complaint behaviors. Weersink and Raymond [4] further demonstrated the influence of education and income on local environmental complaints. Dong et al. [5] demonstrated that exposure to harmful pollutants and household income significantly influence people's complaint behaviors at the provincial level based on economic willingness–to–pay models. Liu [6] verified that the perception of environmental information significantly determined citizens' environmental complaints by questionnaire survey and various multivariate regressions. Tong and Kang [1] explored the relationships between noise complaints and socio–economic factors at the city/region level. Some works indicated social psychological factors that impact environmental complaint behavior on the individual level based on the norm activation model and revealed that the personal norm is the most immediate and powerful predictor of environmental complaint intention [7,8]. Few scholars have discussed the relationship between environmental monitoring data and environmental complaints. Evendijk et al. [9] revealed that hydrocarbons have the highest correlation with the total number of citizen complaints by analyzing the correlation between air measurement results and public complaints. The environmental complaint is one of the most important channels that allows a deeper understanding of the local environment; provides a useful instrument for developing suitable environmental policies; and positively impacts pollution control [10–12]. Arshad et al. [13] constructed an approach to the field of environmental governance by considering youth complaints as an important source of information for the management authorities and verified the effectiveness of the complaint information on environmental governance. Zhang et al. [14] showed that public participation policy plays a significant role in improving environmental governance. A careful review of the existing literature shows that there are limited studies on environmental complaint text mining.

Text mining is the process of extracting previously unknown, understandable, potential, and practical patterns or knowledge from the collection of text data [2]. It has been actively used in various fields, including biomedical, medicine [15], risk management [16], policy, crime [17], market such as multilingual recommendation system [18], education, and informatic fields. Recently, some scholars have carried out research on complaint text. These studies focused on the following aspects: semantic network analysis and keyword analysis of citizen complaints [19]; use of text mining to determine citizens' policy needs for safety and disaster management [20]; and the utilization of text mining to identify and evaluate the indicators of cultural ecosystem services [21]. Overall, previous studies using text mining analysis focus on civil complaints from various viewpoints to provide assistance to the government in decision–making. However, such studies have several limitations: (1) while previous studies are based on civil complaints, few studies have targeted specific urban environmental issues; (2) some only used a certain method of text mining, such as keyword extraction or the semantic network, to analyze the complaint text; therefore, they lacked the systematic application of text mining.

As citizens are direct victims of environmental pollution, the text mining of citizens' complaints will not only help to elucidate their awareness of environmental pollution but also determine more precise countermeasures for the environmental management of smart cities. In this paper, civil environmental complaint records regarding six pollution topics (air, water, noise, waste, electromagnetic radiation, and light) from Guangzhou city are used, and a text mining framework for Chinese environmental pollution complaints is proposed. With this framework, we extract keywords, calculate the complainants' sentiment score, and analyze the characteristics of the semantic network from each class of pollution complaint. These results underline the positive impact of text mining on urban environmental management in both the current and future development of the smart city.

2. Materials and Methods

2.1. Study Area

Guangzhou city is the capital of Guangdong Province, located in the south of mainland China (Figure 1). Guangzhou city is a regional center city in southern China and one of the core cities of the Guangdong–Hong Kong–Macao Greater Bay Area (Greater Bay Area). There are 11 districts in Guangzhou city, and it has a total area of 7434.4 km^2 (2019). At the end of 2019, the resident population of Guangzhou was 15.30 million, and the GDP was RMB 2362.860. According to the list of key polluting firms in Guangzhou city, the number of such firms was 1147, 780, and 713 in 2018, 2019, and 2020, respectively.

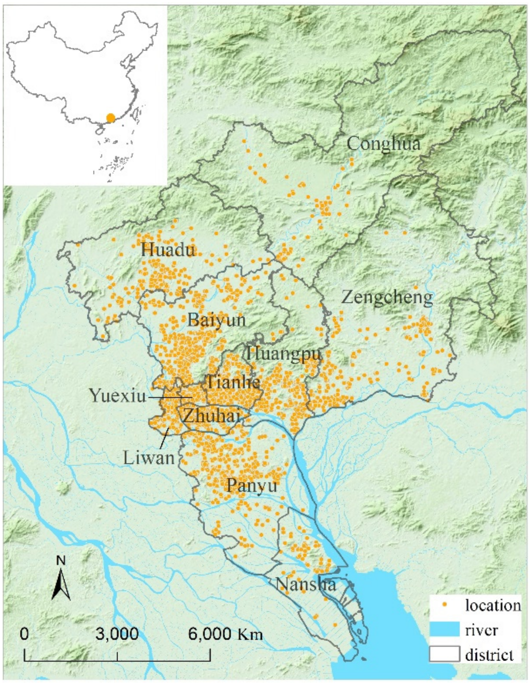


Figure 1. Location of environmental complaints in Guangzhou city (March 2018–March 2020).

2.2. Data Collection and Methods

2.2.1. Data Collection and Pre-Processing

The two–year data (from 1 March 2018 to 31 March 2020) were retrieved from the website of the Guangzhou Municipal Ecological Environment Bureau (http://sthjj.gz.gov.cn/ztlm/tsjbzx/, accessed on 31 March 2020). The complaints datasets contain the date, complaint ID, district and address, firms, topic of complaint, complaint content, government response, and response date (Table 1). We obtained 5672 valid records with missing geographic information, and unidentified complaint content was excluded.

Table 1. A typical example of one complaint record.

Date	29 November 2018 13:03:15
Complaint ID	201811291303154988337
District	黄埔区 Huangpu district
Address	广州经济技术开发区永和经济区田园路西南 Guangzhou Economic and Technological Development Zone, Yonghe Economic Zone Southwest of Tianyuan Road
Firms	广州诺金制药有限公司 Guangzhou Nuojin Pharmaceutical Co., Ltd.
Topic	空气污染 Air pollution
Content	药厂排放废气,严重影响周边环境。 The waste gas emitted by the pharmaceutical factory seriously affects the surrounding environment.
Response	接到投诉后,黄埔区环保局于2018年12月29日到广州诺金制药有限公司现场检查。经查,该公司主要生产中成药,环保手续齐全,在药材炒制、粉碎产生少量粉尘废气和清洗中药废水产生;现场检查时,该公司产生废气经吸尘器处理后高空排放,没有闻到异味。1月25日电话联系投诉人,投诉人表示满意。 After receiving the complaint, the Huangpu District Environmental Protection Bureau conducted an on–site inspection on December 29, 2018. After investigation, the company mainly produces Chinese patent medicines with complete environmental protection procedures. A small amount of dust and waste gas generated during the frying and crushing of medicinal materials and waste water from cleaning Chinese medicine were produced. During on–site inspection, the company’s waste gas was discharged at high altitude after being treated by a vacuum cleaner, and no peculiar smell was smelled. The complainant was contacted by telephone on January 25, and the complainant expressed satisfaction.
Response date	28 January 2019 15:31:25

The 5672 complaint records were classified into six categories, including air, water, noise, waste, electromagnetic radiation (EM radiation), and light based on the topic of complaint (Table 2). Most complaints in all districts regard air pollution, follows by noise, while the categories with the smallest number of complaints are EM radiation and light. The Baiyun district has the largest number of complaints (1174), while the Conghua district has the fewest (157).

Table 2. Records of environmental complaints in each district of Guangzhou.

No.	District	Air	Water	Noise	Waste	EM Radiation	Light	Total
1	Conghua	83	16	54	4	0	0	157
2	Nansha	113	22	47	9	0	1	192
3	Yuexiu	113	8	98	13	3	3	238
4	Liwan	182	33	108	12	3	1	339
5	Zengcheng	254	42	87	4	7	1	395
6	Haizhu	252	34	249	12	0	8	555
7	Huadu	382	54	131	16	4	1	588
8	Huangpu	388	19	223	8	0	5	643
9	Tianhe	313	39	309	15	0	4	680
10	Panyu	402	63	230	14	4	0	713
11	Baiyun	594	127	422	22	5	2	1172
	Total	3076	457	1958	129	26	26	5672

Figure 2 describes the text mining process framework for Chinese environmental complaints. For the sake of content analysis and text mining, we cleaned the collected text data, including removing non–text data (punctuation marks, emoticons, and meaningless symbols), invalid characters (letters and numbers), and meaningless text (function words and pronouns). We removed the meaningless text by using some open–source Chinese stop word dictionaries (e.g., Harbin Institute of Technology (HIT) stop words and Baidu TM stop words). Then, we carried out data processing, including keyword extraction, sentiment analysis, and semantic network analysis.

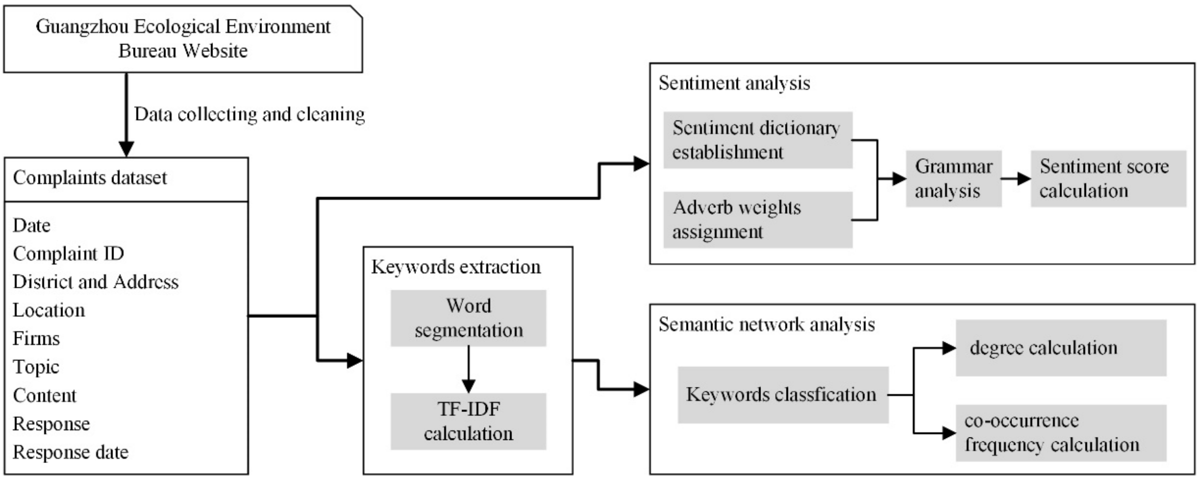


Figure 2. Text mining framework for Chinese environmental complaints.

2.2.2. Keyword Extraction

Firstly, we used the Jieba Chinese text segmentation tool to segment the text records into meaningful words (https://github.com/fxsjy/jieba/, accessed on 25 January 2021). At this stage, synonym substitution and part-of-speech tagging were carried out to avoid the influence of different expressions of synonyms and meaningless function words on subsequent keyword extraction. In addition to the default corpus of the word segmentation tool, a domain dictionary for environmental complaints was established to jointly ensure the accuracy of word segmentation. Secondly, each type of complaint keyword was extracted based on the TF–IDF method [22], which is the most widely adopted word weighting scheme in text mining. It computes how significant a term t is to a document d by combining two scores, term frequency (TF) (2), which is the frequency of term t in document d, and inverse document frequency (IDF) (3), which is the number of documents in the corpus containing t regardless of its frequency. T is more important for d when its TF is large but its IDF is small. That is, words with high TF-IDF value are more important than other words in the documents, so they are the keywords that distinguish the document from others.

$$\mathrm{TF}-\mathrm{IDF}=\mathrm{TF}\times\mathrm{IDF} \tag{1}$$

$$\mathrm{TF}=\frac{f(t,\mathrm{d})}{|d|} \tag{2}$$

$$\mathrm{IDF}=\log\frac{|D|}{|\{d|t\in d\}|} \tag{3}$$

where f (t, d) is the number of times term t appears in a document, d is the total number of terms in the document, D is the total number of documents, and $|\{d|t\in d\}|$ is the number of documents with the term t in it.

2.2.3. Sentiment Analysis

In this study, sentiment analysis was used to identify the citizen's sentiment in the six types of environmental complaints. Lacking inter–word spacing, the diversification of expressions, the complexity of grammar, and the randomness of length of the complaint record increase the difficulty of Chinese sentiment analysis.

Firstly, a sentiment dictionary was established, including a domain emotion dictionary of environmental complaints and some general Chinese sentiment dictionaries, such as Li Jun's Chinese commendatory and derogatory dictionary of Tsinghua University, National Taiwan University Sentiment Dictionary (NTUSD), Hownet Sentiment Dictionary. Meanwhile, the score of positive emotion words (Sp) was set to 1, and the score of negative emotion words (Sn) was -1 (Table 3).

Table 3. Sentiment words and their weights.

Lexicon	Examples of Sentiment Words	Emotion	Weight
General	开心 (happy), 公平 (fair), 心爱 (beloved)	Positive	1
	不幸 (unfortunate), 狂怒 (furious), 狠心 (heartless)	Negative	−1
Domain	安全 (safety), 干净 (clean), 舒服 (comfortable)	Positive	1
	危害 (harmful), 刺激 (irritation), 刺耳 (piercing)	Negative	−1

Secondly, according to Hownet Dictionary, degree adverbs are divided into six levels. According to the weight value of the gradient descent Formula (4) [23], different weights are assigned to each level (Table 4). The emotional intensity of the emotional words modified by adverbs increases by a certain multiple. Moreover, when inverse words such as scarcely (没有), never (从不), and seldom (很少), modify emotional words, the emotional words are multiplied by -1.

$$Aw_{n+1} = A_w\left(\frac{\sqrt{2}}{2}\right)^n,\ n = 1,2,3,4,5 \tag{4}$$

where, $A_w = 3$ is the weight of the "most" level; $\left(\frac{\sqrt{2}}{2}\right)^n$ is the gradient descent rate.

Table 4. Degree adverbs and its weights.

Level	Examples of Adverb (A) and Inverse Words (N)	Weight (Aw)
Most	超级 (super), 极其 (extremely), 最 (most)	3
Very	特别 (special), 非常 (very), 尤其 (especially)	2.1
More	更 (more), 较 (relatively), 越是 (more)	1.5
Ish	略微 (slightly), 一些 (some), 有点 (a little)	1.06
Insufficiently	仅仅 (merely), 不太 (not too), 相对 (relative)	0.75
Over	不为过 (not too much), 略多 (slightly more)	0.53

Finally, one complaint record (a compound sentence) is divided into multiple clauses by punctuation, and the sentiment value of each clause (Ci) is calculated by the combination of sentiment words (S), adverbs (A), inverse words (N), and punctuation (!/?) (Table 5). Additionally, the sentiment value of each complaint record (Sj) is calculated by Function (5). Table 5 shows nine combinations in Chinese grammar.

$$S_j = \frac{\frac{\sum_{i=1}^{n}(Ci)}{L_j}}{\left|\max\left(S_j\right)\right|} \tag{5}$$

where S_j is the sentiment value of the j complaint record, L_j is the clauses' number of j complaint records, and Ci is the sentiment value of the i clause in the j complaint record.

L_j is used to eliminate the influence of the complaint record's length on the result. The sentiment value (S_j) is scaled in the range -1–1. $S_j > 0$ means the sentiment of the complaint is positive; $S_j < 0$ means the sentiment is negative; $S_j = 0$ means the sentiment is neutral.

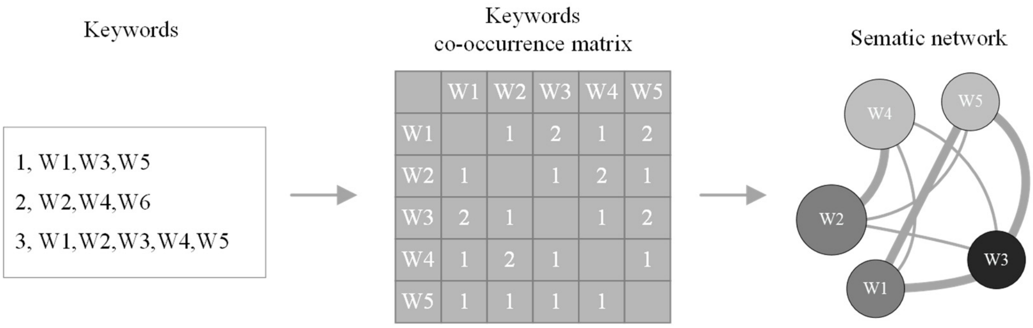
Table 5. Common combinations of compound sentences.

No.	Combination	Example	*Ci*	Score
1	S	开心 (happy)	Sp	1
2	S + !/?	开心! (happy!/happy?)	Sp + 2/−2	3/−1
3	N + S	不开心 (not happy)	(−1) × Sp	−1
4	N + N + S	不是不开心 (not unhappy)	Sp	1
5	N + A + S	不是非常开心 (not very happy)	0.5 × Aw × Sp	1.1
6	A + S	非常开心(very happy)	Aw × Sp	2.1
7	A +A + S	非常非常开心 (very, very happy)	(Aw + Aw) × Sp	4.2
8	A + N + S	非常不高兴 (very unhappy)	1.5 × (−1) × Aw × Sp	−3.15
9	S + A	危害极大 (extremely harmful)	Aw × Sn	−3

2.2.4. Semantic Network Analysis

A semantic network consists of nodes (words) and edges (the relationship between words). The node's size (degree) is proportional to the number of words related to it; a thicker edge means a higher co–occurrence frequency or a closer relationship between the words. We used two–mode networks [24], including top and bottom nodes, to analyze the semantic network of each type of complaint. In our two–mode networks, keywords (bottom nodes) were categorized into three clusters (top nodes) based on pollution characteristics, stakeholders, or complainants. Furthermore, the pollution characteristics were categorized into three sub–clusters including pollution sources, pollution behavior, and sensory features; the stakeholders were categorized into two sub–clusters, including firms and administration; and the complainants were categorized into three sub–clusters, including pollution receptor, social life, and individual health.

Figure 3 shows the workflow of semantic network analysis. Firstly, keywords were extracted based on the TF–IDF method. Secondly, a word co–occurrence matrix with environmental complaint keywords was constructed, and co–occurrence analysis was performed on them. Finally, the generated semantic network was plotted by Gephi software (version 0.9.2) [25].



1, W1,W3,W5
2, W2,W4,W6
3, W1,W2,W3,W4,W5

	W1	W2	W3	W4	W5
W1		1	2	1	2
W2	1		1	2	1
W3	2	1		1	2
W4	1	2	1		1
W5	1	1	1	1	

Figure 3. The workflow of semantic network analysis.

3. Results and Discussion

3.1. Keywords of Environmental Complaints

The study used TF–IDF to extract keywords from six types of environmental complaints that indicated the characteristics of environmental complaints. The higher the TF-IDF value, the more important the word is in this type of environmental complaint. Table 6 shows the top 10 keywords of various environmental complaints, and we found that different environmental complaints show obvious differences and similarities characteristics of environmental issues.

Table 6. Top 10 keywords of environmental complaints and their TF–IDF value.

Air			Water			Noise			Waste			EM Radiation			Light		
Keyword		TF–IDF	Keyword		TF–IDF	Keyword		TF–IDF	Keyword		TF–IDF	Keyword		TF–IDF	Keyword		TF–IDF
居民	resident	149.75	污水	sewage	33.50	噪音	noise	180.32	垃圾	waste	13.54	换流站	converter station	2.94	小区	community	2.26
油烟	lampblack	138.85	居民	resident	19.55	居民	resident	109.49	清理	clean up	7.23	项目	project	2.82	居民	resident	2.16
废气	exhaust	122.62	恶臭	stench	12.52	扰民	disturb	87.16	小区	community	6.46	信号	signal	2.72	外墙	exterior wall	1.79
气味	odor	120.87	工厂	factory	11.79	声音	sound	52.09	居民	resident	5.72	基站	base station	2.69	严重	serious	1.66
工厂	factory	97.01	环境	surrounding	11.70	小区	community	47.80	环境	surrounding	5.63	居民	resident	2.19	通宵	overnight	1.54
小区	community	94.48	村民	villager	11.67	分贝	decibel	44.72	建筑	building	5.51	电磁辐射	electromagnetic radiation	1.88	射灯	spotlight	1.49
部门	department	82.03	部门	department	11.10	部门	department	44.40	垃圾桶	ashbin	5.21	规划	planning	1.84	强光	glare	1.35
健康	health	79.79	下水道	sewer	11.07	噪声	noise	44.17	村民	villager	4.68	楼顶	roof	1.80	广告牌	billboard	1.15
味道	smell	78.99	气味	odor	9.83	油烟	lampblack	42.63	部门	department	4.51	屋主	homeowner	1.79	扰民	disturb	1.09
垃圾	waste	75.79	废气	exhaust	9.77	粉尘	dust	37.97	土壤	soil	4.29	距离	distance	1.62	平台	platform	1.06

As the keyword list demonstrates, differences in environmental complaints with different topics are noticeable. The list of keywords related to air complaints has the highest TF–IDF value for typical words, such as lampblack (油烟), exhaust gas (废气), and odor (气味). Among the keywords of water complaints, sewage (污水) ranks first, followed by stench (恶臭), sewer (下水道), and smell (气味). In noise complaints, the most important word is noise (噪音), followed by sound (声音) and decibel (分贝) also showing high scores. The word with the highest TF–IDF value in the waste complaint is waste (垃圾), which also includes feature words, such as waste cleaning (清理) and ashbin (垃圾桶). The most critical vocabulary in EM radiation complaints consists of converter station (换流站), signal (信号), base station (基站), and EM radiation (电磁辐射). The keywords for light complaints are community (小区) and resident (居民).

In short, this proves that keywords can accurately reflect the differences in environmental complaints and further provide a scientific basis on which for environmental managers to solve environmental problems with accurate entry points. Turning to the similarities of keywords, the terms resident (居民) and community (小区) appear in all type of complaints. The result confirms that the residents and their living environment are of great concern in environmental complaints.

3.2. The Sentiment of Environmental Complaints

The box plot (Figure 4) shows that the mean (air: −0.11; water: −0.10; noise: −0.10; waste: −0.04; EM radiation: −0.15; light: −0.18) and median (air: −0.09; water: −0.08; noise: −0.08; waste: −0.04; EM radiation: −0.10; light: −0.19) of all types of environmental complaint sentiment are both lower than zero, which indicates that the complainants' overall sentiment tendency is negative. Comparing the mean and median of various environmental complaints, electromagnetic radiation and light have the lowest value. The sentiment value distribution of electromagnetic radiation is the most scattered (0.30), followed by light (0.23), which is presumably due to the wide differences between cognitive and individual. There is little difference in the sentiment value distribution of air, water, and noise pollution complaints.

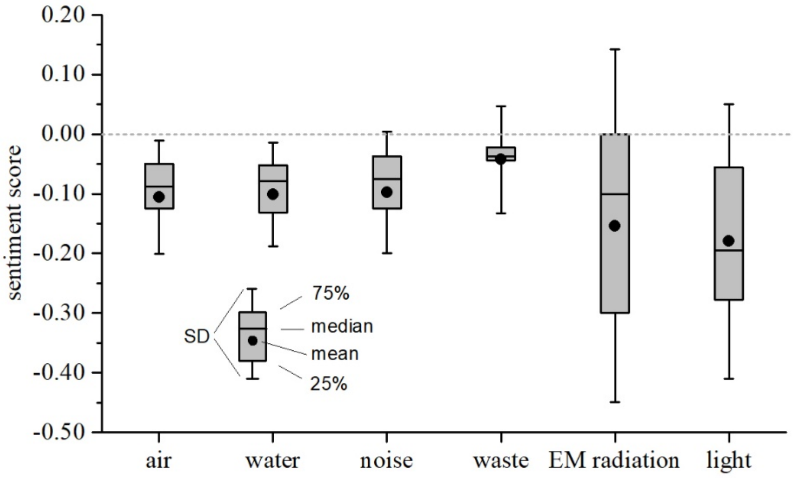


Figure 4. Sentiment score for each complaint.

3.3. The Semantic Network of Environmental Complaints

As shown in Table 7, we identified the proportion of clusters and sub–clusters in semantic networks. From the semantic network node, the pollution characteristic is the largest cluster of each network. Except for noise complaints, cluster 3 (complainant) has a higher proportion than cluster 2 (stakeholder). This suggests that individuals making the complainants pay most attention to pollution characteristics, especially the sub–cluster pollution source, followed by their impacts. Stakeholders account for the smallest proportion, which may indicate the least understanding of this cluster of complainants.

Table 7. Statistics of semantic network clusters of each complaint.

Cluster	Sub–Cluster	Air	Water	Noise	Waste	EMR	Light
1. Pollution characteristic	Pollution source (PS)	29.17%	49.45%	49.15%	38.62%	38.03%	54.65%
	Pollution behavior (PB),	13.54%	8.79%	11.32%	11.88%	14.08%	13.95%
	sensory features (SF)	15.62%	10.99%	5.65%	5.94%	5.63%	3.49%
2. Stakeholder	Firms (FM),	10.42%	6.59%	11.86%	11.88%	2.82%	2.33%
	administration (AD)	7.29%	4.39%	4.52%	2.97%	5.63%	0
3. Complainant	Pollution receptor (PR),	11.46%	12.09%	8.47%	16.83%	9.86%	9.3%
	social life (SL),	6.25%	4.4%	7.34%	8.91%	11.27%	9.3%
	individual health (HL)	6.25%	3.3%	1.69%	2.97%	12.68%	6.98%

Citizens' insufficient knowledge of relevant stakeholders, such as polluting firms and administrations, has also led to complaints that cannot be handled well. According to the official statistics of responses to complaints, 1225 complaints (21.60%) are not within the authority of the Ecology Environment Bureau. Moreover, the complaint contained other stakeholders, including the Water Affairs Bureau, the Urban Management Bureau, and the Education Bureau, which reflects the complexity of urban pollution management. Therefore, urban environmental management needs to strengthen the coordination of multiple departments.

Figure 5 reflects the relationships between the keywords of citizens' environmental complaints, from which we observed that the relationships between pollution sources and pollution receptors (PR–PS) are the most important in environmental complaints, such as resident–lampblack (居民–油烟) and resident–exhaust gas (居民–废气) in air complaint; resident–sewage (居民–污水) and residential–oil bath (住宅–油池) in water pollution complaints; noise–resident (噪声–居民) and resident–lampblack (居民–油烟) in noise complaints; waste–resident (垃圾–居民) and garbage station–resident (垃圾站–居民) in waste complaints, residential–converter station (住宅–换流站) in electromagnetic radiation complaints; and LED–resident (LED–居民) in light pollution complaint. From the standpoint of the complainant, pollution sources are a primary concern in environmental complaints. The relationships between the above keywords indicate which pollution should be first supervised and controlled.

In addition to the most concerning relationship between pollution sources, other relationships in environmental complaints also deserve the attention of environmental managers, including those between pollution receptors and pollution behavior (PR–PB), pollution receptors and sensory feature (PR–SF), and pollution receptors and individual health (PS–HL) (Table 8). As shown in Figure 5, complaints about pollution behavior (PB) mostly regard space and time. The pollution behavior of air complaints and waste complaints emphasizes spatial issues (people–location '人民–选址' and resident–location '居民–选址'), while the pollution behavior of noise complaints and light complaints emphasizes time, such as resident–disturbing (居民–扰民), residential–disturbing (住宅–扰民), and resident–overnight (居民–通宵). The relationship between the pollution receptor and sensory feature (PR–SF) is more prominent in air and waste complaints, mainly for smell–related terms, such as residential and odors (住宅–气味) and resident and stench (居民–臭味). Complaints about EM radiation show that the relationship between pollution receptors and individual health (PR–HL) is more prominent. Specifically, citizens are most concerned about the impact of converter stations on safety and health (converter station–physical and mental health 换流站–身心健康). This suggests that supervisors should provide the public with EM radiation–related knowledge.

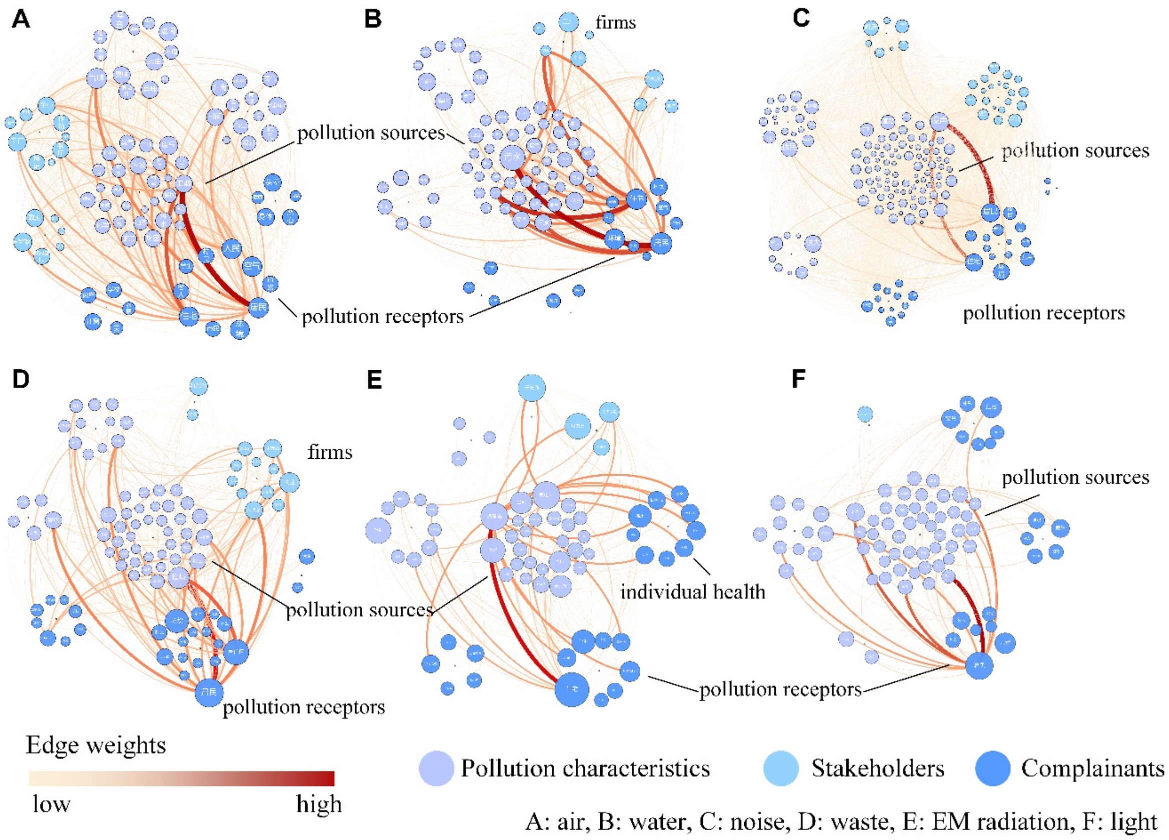


Figure 5. The semantic network of environmental complaints. (**A**): 96 nodes and 1371 edges; (**B**): 91 nodes and 582 edges; (**C**): 177 nodes and 2683 edges; (**D**): 101 nodes and 458 edges; €: 72 nodes and 302 edges; (**F**): 86 nodes and 252 edges.

Table 8. Top 10 relations of environmental complaints semantic networks.

	Air			Water			Noise	
Relation	**Edge**	**Weight**	**Relation**	**Edge**	**Weight**	**Relation**	**Edge**	**Weight**
PR–PS	居民–油烟 resident–lampblack	1196	PR–PS	居民–污水 resident–sewage	114	PS–PR	噪声–居民 noise–resident	1255
PR–PS	住宅–油烟 residential–lampblack	849	PR–PS	住宅–油池 residential–oil bath	100	PR–PS	住宅–噪声 residential–noise	868
PR–SF	居民–气味 resident–smell	647	PR–PS	居民–河流 resident–river	83	PR–PS	居民–油烟 resident–lampblack	456
PR–PS	居民–废气 resident–exhaust gas	596	PR–FM	住宅–商场 residential–mall	80	PR–PS	住宅–油烟 residential–lampblack	422
PR–PS	人民–垃圾 people–waste	512	PR–PS	居民–油池 resident–oil bath	79	PS–PR	噪声–环境 noise-environment	339
PR–PS	住宅–废气 residential–exhaust gas	507	PR–PS	住宅–垃圾 residential–waste	71	PB–PR	很大–居民 very noisy-resident	268
PR–AD	居民–环保局 resident–Environmental Protection Agency	483	PR–PS	住宅–污水 residential–sewage	66	PR–PB	居民–扰民 resident–disturb	253
PR–PB	人民–选址 people–location	480	FM–PR	商场–居民 mall–resident	64	PS–PS	噪声–道路 noise–road	243

Table 8. *Cont.*

PR–PS	住宅–垃圾 residential–waste	478	PR–PS	住宅–广场 residential–square	60	PS–AD	噪声–政府 noise–government	200
PR–SF	住宅–气味 residential–smell	444	FM–PS	商场–油池 mall–oil bath	60	PR–PB	住宅–扰民 residential–disturb	194
	Waste			**EM radiation**			**Light**	
Relation	**Edge**	**Weight**	**Relation**	**Edge**	**Weight**	**Relation**	**Edge**	**Weight**
PS–PR	垃圾–居民 waste–resident	55	PR–PS	住宅–换流站 residential-converter station	64	PS–PR	LED–居民 LED–resident	17
PS–PR	垃圾–住宅 waste–residential	38	PS–FM	变电站–开发商 substation-developer	32	PR–PS	居民–灯光 resident–light	12
PS–PR	垃圾站–居民 garbage station–resident	31	PS–HL	变电站–安全 substation-safety	32	PR–PS	居民–楼盘 resident–real estate	12
PR–SF	居民–臭味 resident–stench	30	PS–HL	换流站–身心健康 converter station-physical and mental health	31	PB–PR	刺眼–居民 glare-resident	10
PR–PB	居民–选址 resident–location	28	HL–PS	健康–换流站 health-converter station	29	PB–PR	光污染–居民 light pollution-resident	8
PR–PS	居民–蚊虫 resident–mosquito	25	PR–PS	居住环境–换流站 living environment-converter station	29	PR–PB	居民–施工 resident-construction	8
PS–PR	垃圾–环境 waste-environment	25	PR–PS	儿童–换流站 children-converter station	28	PR–PB	居民–通宵 resident-overnight	7
PS–PR	垃圾桶–住宅 ashbin–residential	25	PS–AD	换流站–电力局 converter station-power bureau	28	PS–PR	射灯–居民 spotlights-resident	7
PR–PS	住宅–蚊虫 residential–mosquito	23	PS–PR	换流站–聚居区 converter station-residential area	28	PS–PR	噪音–居民 noise–resident	6
PR–FM	居民–物业 resident–property	22	PS–HL	换流站–死亡率 converter station-mortality rate	28	PS–SL	平台–生活 platform-life	6

The relationship between pollution receptors and pollution behavior (PR–PB) suggests that scientific and integrated site selection is necessary to resolve environmental complaints, including more reasonable site selection of garbage dumps and power telecommunication equipment and stricter construction time control measures. Actions should be taken to address the problems reflected by sensory features (such as stench, mosquitoes, and rats) and to provide the public with environmental and scientific knowledge, especially regarding EM radiation pollution.

4. Conclusions

In this study, a framework for the textual analysis of Chinese environmental protection complaints was established, and the two–year civil environmental complaint records in Guangzhou city were analyzed using this framework. The conclusions show the following: (1) Civil environmental complaint characteristics can be identified. Keywords of various types of environmental complaints can be automatically and effectively extracted by TF–IDF, such as "lampblack" and "exhaust gas" in air pollution and "LED lights" in light pollution, which provides an accurate entry point for solving urban environmental problems. It also provides technical support for smart city environmental management. (2) The overall sentiment of environmental complaints is negative. Light pollution complaints are the most negative, and EM radiation complaints have the most fluctuating emotions, which may be caused by differences in citizen perception of EM radiation. (3) The semantic network nodes of the six types of environmental complaints reveal that the public pays the most attention to the pollution sources when complaining but the least attention to stakeholders, which may reduce the efficiency of environmental managers in handling complaints. (4) Besides the Ecology Environment Bureau, stakeholders in environmental complaints involve multiple government departments, including water affairs departments, urban management departments, and other departments. This not only reflects the complexity of environmental pollution but also shows that the issue of environmental complaints is deemed urgent by multiple departments. (5) The citizen semantic network indicates that pollution sources and pollution receptors are paid the most attention. Simultaneously, among different types of complaints, the pollution receptor's relationship with pollution behaviors (site selection, overnight construction), sensory features (stench, dazzle), stakeholders, and individual health are also highlighted by citizens. These relationships suggest that the pollution behavior of pollution sources, sensory features, environmental knowledge of pollution sources, and other details may become a crucial part of pollution management, which will provide more accurate management measures and be beneficial to smart urban environmental governance.

For accurate text mining in further research, a rich corpus of environmental complaints must be established, and adaptable Chinese grammar for complaints needs to be summarized. Named–entity recognition could be considered, which will provide assistance in extracting detailed information about pollution incidents in semantic network analysis. Urban environmental management departments must establish a big data analysis system for environmental complaints based on text mining technology. Only in this way can urban environmental issues be effectively managed.

Author Contributions: Y.J. developed the framework for textual analysis and performed the experiments, derived the models, and analyzed the data. Y.L. was involved in part of the code work. Y.J. wrote the manuscript in consultation with C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences [Grant No. XDA23030401] and the Youth Innovation Promotion Association, Chinese Academy of Sciences [Grant No. 2017351].

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tong, H.; Kang, J. Relationships between noise complaints and socio-economic factors in England. Sustain. *Cities Soc.* **2021**, *65*, 102573. [CrossRef]
2. Zhang, Y.; Chen, M.; Liu, L. A review on text mining. In Proceedings of the 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 23–25 September 2015; pp. 681–685.
3. Dasgupta, S.; Wheeler, D. Citizen Complaints as Environmental Indicators: Evidence from China. In *The Causal Effects of Long-Term PM2.5 Exposure on COVID-19 in India*; The World Bank: Washington, DC, USA, 1997.
4. Weersink, A.; Raymond, M. Environmental regulations impact on agricultural spills and citizen complaints. *Ecol. Econ.* **2007**, *60*, 654–660. [CrossRef]
5. Dong, Y.; Ishikawa, M.; Liu, X.; Hamori, S. The determinants of citizen complaints on environmental pollution: An empirical study from China. *J. Clean. Prod.* **2011**, *19*, 1306–1314. [CrossRef]
6. Liu, X.; Dong, Y.; Wang, C.; Shishime, T. Citizen Complaints about Environmental Pollution: A Survey Study in Suzhou, China. *J. Curr. Chin. Aff.* **2011**, *40*, 193–219. [CrossRef]
7. Zhang, X.; Geng, G.; Sun, P. Determinants and implications of citizens' environmental complaint in China: Integrating theory of planned behavior and norm activation model. *J. Clean. Prod.* **2017**, *166*, 148–156. [CrossRef]

8. Zhang, X.; Liu, J.; Zhao, K. Antecedents of citizens' environmental complaint intention in China: An empirical study based on norm activation model. *Resour. Conserv. Recycl.* **2018**, *134*, 121–128. [CrossRef]
9. Evendijk, J.; Müskens, P.; De Jong, T. Relationship Between Citizen Complaints of Air Pollution, Meteorological Data and Immission Concentrations. *Stud. Environ. Sci.* **1980**, *8*, 379–386. [CrossRef]
10. Huang, H.; Miller, G.Y. Citizen Complaints, Regulatory Violations, and Their Implications for Swine Operations in Illinois. *Appl. Econ. Perspect. Policy* **2006**, *28*, 89–110. [CrossRef]
11. Carvalho, D.S.; Fidélis, T. The perception of environmental quality in Aveiro, Portugal: A study of complaints on environmental issues submitted to the City Council. *Local Environ.* **2009**, *14*, 939–961. [CrossRef]
12. Wang, H.; Di, W. The Determinants of Government Environmental Performance: An Empirical Analysis of Chinese Townships. In *The Causal Effects of Long-Term PM2.5 Exposure on COVID-19 in India*; The World Bank: Washington, DC, USA, 2002; pp. 704–708.
13. Arshad, S.; Shafqat, A.; Khan, A.A.; Safdar, Q. Youth environmental complaints in Bahawalpur City, Pakistan: An informational intervention for local environmental governance. *Hum. Geogr. J. Stud. Res. Hum. Geogr.* **2013**, *7*, 71–80. [CrossRef]
14. Zhang, G.; Deng, N.; Mou, H.; Zhang, Z.G.; Chen, X. The impact of the policy and behavior of public participation on environmental governance performance: Empirical analysis based on provincial panel data in China. *Energy Policy* **2019**, *129*, 1347–1354. [CrossRef]
15. Bhasuran, B.; Subramanian, D.; Natarajan, J. Text mining and network analysis to find functional associations of genes in high altitude diseases. *Comput. Biol. Chem.* **2018**, *75*, 101–110. [CrossRef]
16. Jacinto, R.; Reis, E.; Ferrão, J. Indicators for the assessment of social resilience in flood-affected communities—A text mining-based methodology. *Sci. Total Environ.* **2020**, *744*, 140973. [CrossRef]
17. Tseng, Y.H.; Ho, Z.P.; Yang, K.S.; Chen, C.C. Mining term networks from text collections for crime investigation. *Expert Syst. Appl.* **2012**, *39*, 10082–10090. [CrossRef]
18. Liu, P.; Zhang, L.; Gulla, J.A. Multilingual Review-aware Deep Recommender System via Aspect-based Sentiment Analysis. *ACM Trans. Inf. Syst.* **2021**, *39*, 1–33. [CrossRef]
19. Min, K.; Jun, B.; Lee, J.; Kim, H.; Furuya, K. Analysis of Environmental Issues with an Application of Civil Complaints: The Case of Shiheung City, Republic of Korea. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1018. [CrossRef]
20. Lee, E.; Lee, S.; Kim, K.S.; Pham, V.H.; Sul, J. Analysis of Public Complaints to Identify Priority Policy Areas: Evidence from a Satellite City around Seoul. *Sustainability* **2019**, *11*, 6140. [CrossRef]
21. Lee, J.-H.; Park, H.-J.; Kim, I.; Kwon, H.-S. Analysis of cultural ecosystem services using text mining of residents' opinions. *Ecol. Indic.* **2020**, *115*, 106368. [CrossRef]
22. Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manag.* **1988**, *24*, 513–523. [CrossRef]
23. Xin, Y.; Yang, Y.; Jiao, W.; Zhu, D.; Zheng, S.; Yuan, Z.; Yang, X.; Luo, Z. Sentiment Analysis of Homestay Comments Based on Domain Dictionary. *Sci. Technol. Eng.* **2020**, *020*, 2794–2800.
24. Opsahl, T. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Soc. Netw.* **2013**, *35*, 159–167. [CrossRef]
25. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In Proceedings of the Third International Conference on Weblogs and Social Media, San Jose, CA, USA, 17–20 May 2009.