Public Concern at International Sporting Events: Analysis of Hotspots and Expression Patterns in Complaint Texts

Nannan Liang[1,a], Zan Chen[2,b], Jieming Bai[3,c]

{liangnanan@zhejianglab.com[a], chenz@zhejianglab.com[b], baijm@zhejianglab.com[c]}

Research Center for Data Hub and Security, Zhejiang Lab, Hangzhou, China[1]
Research Center for Development Strategy and Cooperation Center, Zhejiang Lab, Hangzhou, China [2, 3]

Abstract. Prompt resolution of public demands contributes to a stable social environment for the smooth running of offline international sporting events. To gain a comprehensive understanding of the public concerns within international sporting events, this study utilizes online complaint data related to the Asian Games from the "Min Hu Wo Wei" platform in Zhejiang Province, China by applying artificial intelligence techniques such as large language models and BERTopic framework, in conjunction with domain knowledge-driven theoretical reasoning methods. We have developed a systemic analytical framework that combines "data and domain knowledge. The findings reveal that the most prevalent category of complaints related to the Asian Games is "urban planning and public facilities," while public grievances tend to concentrate on issues indirectly associated with the sporting event. Furthermore, this study further explores the hotspots of public complaints in the context of the "Asian Games", and utilizing a two-dimensional quadrant based on "publicity and relevance" to analyze the complaint regularity through expression patterns, and proposes corresponding governance strategies.

Keywords: International sporting events; social governance; livelihood issues; large language models; text analysis

1 Introduction

After the pandemic, international exchanges have become frequent, and cities have hosted offline international events. In 2023, China hosted consecutive major sports events such as the 31st World University Games in Chengdu and the 19th Asian Games in Hangzhou, the planning process required the participation of the entire urban system. During this period, varying degrees of overlap occurred between urban construction and renovation, such as infrastructure, road planning, and venue construction, and the lives and work of the people. Therefore, hosting international events constitutes both an a test of the urban construction and a challenge to governance capacity. In addition to paying attention to the urban branding effects of international sporting events, governments should also endeavor to ensure the citizen's livelihood needs and alleviate social tensions in time to prevent governance failures arising from livelihood issues.

An essential prerequisite for achieving the aforementioned goals is to comprehend the impact of international sporting events on decentralized and complex livelihood issues. Existing social

PMIS 2024, March 15-17, Changsha, People's Republic of China

DOI 10.4108/eai.15-3-2024.2346421

science research typically adopts three conventional methods to investigate livelihood issues. Firstly, field surveys and case studies are conducted to generate a series of research material. This method primarily relies on textual records and often involves in-depth investigations of specific social events[1]. Secondly, scholars collect structured data from yearbooks, reports, or questionnaire surveys for quantitative analysis. This method is typically used to identify key factors related to livelihood issues [2]. Thirdly, scholars employ inference based on experiential knowledge. After accumulating substantial experiential knowledge in the fields of public administration and social governance, scholars extend theoretical frameworks to address specific livelihood issues [3].

Previous research heavily relied on researchers' own capabilities on material collection and variable measurement, primarily reflecting long-term cumulative social issues. With the advent of online channels for public expression and feedback, the avenues for public demands and opinions have expanded. A vast amount of real-time governmental website and online complaint data provide scholars with new insights into understanding livelihood demands and discovering their underlying mechanisms [4].

Existing research can be classified into two types: one focuses on optimizing text classification techniques to improve the accuracy of identifying complaint text topics. [5] apply BERT combined with the LinearSVC algorithm to classify public political inquiry records into seven categories. [6], [7] employed LDA topic modeling to explore the distribution patterns of complaints received by a traffic service hotline and quickly identify the topic distribution of short texts within a short time. [8] used Latent Dirichlet Allocation (LDA) to extract latent topics from complaint narratives in the Consumer Financial Protection Bureau (CFPB) dataset. [9] proposed an integrated method that combines a dual-tower recommendation model and BERTGraph to recommend historical similar complaints for unresolved complaints, assisting hotline operators in more accurately allocating complaints. [10] utilized the BERT-GCN method to extract text features and then performed text classification tasks using a multi-attention mechanism.

The other type of research builds upon text classification to explore the support of complaint data for specific governance tasks. Some scholars have utilized complaint text descriptions of noise pollution [11], air quality [12], and other environmental issues, to supervise and track environmental problems. They have also supplemented spatial distribution information in urban management through public complaints about street vendors' irregular operations [13]. Furthermore, by analyzing public complaints related to landslide incidents [14] and outbreaks of foodborne diseases [15], they have assisted governments in predicting and managing crisis.

Overall, traditional research on livelihood issues has high requirements for structured data and heavily relies on the extensive expertise of researchers. Although some advanced studies have introduced data analysis methods, such as natural language processing, to handle unstructured complaint text data, to which most of the research related still focuses primarily on parameter optimization. Due to the insufficient integration of "data" and "domain knowledge" in social governance, the current research lacks in-depth analysis on mining hot topics of public concern and patterns of demands expression. In response to this, this study integrates artificial intelligence technologies including large language models and BERTopic with traditional theory-based reasoning methods to explore public concerns in international sporting events.

This study contributes in the following aspects: (1) It conducts in-depth analysis on the types of public issues, mining hot topics, and expression patterns during major international public events based on citizen complaint data. This study fills the gap in government governance research by enhancing the understanding of public sentiments and concerns during international sporting events. (2) It designs and implements an analysis framework that integrates of "data" and "domain knowledge", which breaks through the one-sided research approach to livelihood issues in the field of social governance, expanding the scope of research in intelligent social governance.

2 Research design

2.1 Data preparation

This study collected public complaint text from the unified platform "Min Hu Wo Wei" on the Zhejiang Provincial Portal. The platform integrates the handling of counselling, appeals, suggestions, and other matters across various departments, prefectures, and municipalities. The data search span ranges from 16 September 2015, when Hangzhou successfully bid to host the 19th Asian Games, to 31 October 2023, the month in which the Games were held.

The search rule involves identifying matters containing the keyword "Asian Games" in the "Title/Content" section of the "Disclosure of matters" webpage on the unified platform. Consultations and suggestions were excluded from the search result and only complaints were retained. A total of 4,912 online complaints were obtained, with the earliest dating back to 21 July 2017. The metadata of the complaint data includes the title, content description, and registration time. We used the most informative complaint content as input data for the entire research process.

2.2 Research method

2.2.1 Similarity-based text classification

Similarity-based text classification consists of two main steps: text embedding and similarity calculation.

This paper utilizes the M3e pre-trained language model based on the BERT framework to characterize Chinese semantic information more precisely. BERT is a pre-trained language modeling framework based on the Transformer architecture (Fig.1). It generates contextually relevant vector representations for each word by utilizing bi-directional contextual information. This allows it to be employed in various natural languaged processing tasks via transfer learning, such as text classification, named entity recognition, etc [16]. In Chinese scenarios, the M3e model exhibits higher accuracy. M3e is a pre-trained language model based on the BERT framework developed by the MokaHR team. The model was trained on more than 22 million Chinese sentence pair dataset with 1 epoch, using A100 80G maximized batch-size to ensure negative sampling. A experiment indicates that M3e outperforms text2vec and text-embedding-ada-002 on average on Chinese homogeneous text datasets and surpass both on Chinese retrieval tasks [17] .

Cosine similarity is simple yet efficient algorithm commonly used for calculating text similarity. This method does not necessitate a large amount of labeled data, it operates as a form of unsupervised learning that rely solely on textual descriptions of categories. Therefore, the cosine similarity calculation method is adopted in this research.

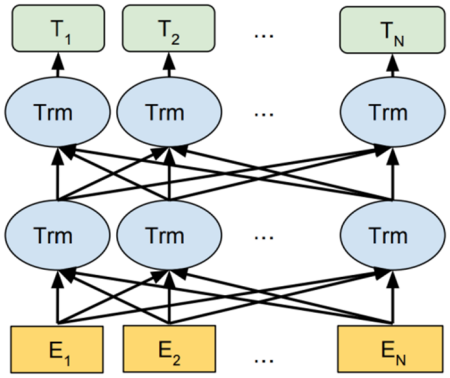

Fig. 1. Architecture of the BERT model

2.2.2 LLM based named entities recognion

The Large Language Model (LLM) is a deep learning model trained on an extensive corpus text data. It is capable of capturing the semantic of the text while also generating natural language text. The 2023 Baichuan2-13B LLM, developed by Baichuan Intelligence, is a new generation of open-source LLM tailored for Chinese language scenarios. It is trained with a high-quality corpus of 2.6 trillion tokens. The Baichuan2-13B LLM outpeforms models of the same size on multiple authoritative benchmarks for Chinese, English and multilingual general-purpose and domain applications. Therefore, in this paper, we employ it to perform Named Entities Recognion on complaint texts for subsequent hotspot analysis of Chinese complaints [18].

2.2.3 Hotspot identification technique based on BERTopic framework

BERTopic is a topic modeling framework based on the BERT model. It automatically identifies structures in large-scale text by learning semantic representations via pre-trained language models. The BERTopic framework consists of five steps: Embedding, Dimensionality reduction, Clustering, Segmentation Tokenizer, and Weighting scheme. These steps create clusters and present thematic information of the clustered texts to achieve hotspot recognition within each cluster [19].

2.3 Technological roadmap

2.3.1 Text classification

The paper uses the M3e-base model to separately embed complaint and category label as shown in Fig.2a, yielding 768-dimensional vectors for each. The cosine similarity algorithm is used to calculate the similarity between each complaint content vector and each category label vector, and the complaint content is then assigned to the category with the highest similarity. If

the similarity of any category is lower than the thereshold, it will be categorized as "Others". This categorization process applies to all complaints.

2.3.2 Extraction of named entities from complaint content

Fig.2b depicts the Named Entity Recognition process by Baichuan2-13B LLM. We combine the specific complaint content with the appropriate prompt, and call the API of the Baichuan2-13B LLM, and then extract the key entities from each complaint content (the object of the complaint, the place of the incident, the reason for the complaint, the complainant's demand, etc.) through its ability of text understanding and generation. This step aims to remove invalid information in the complaint content.

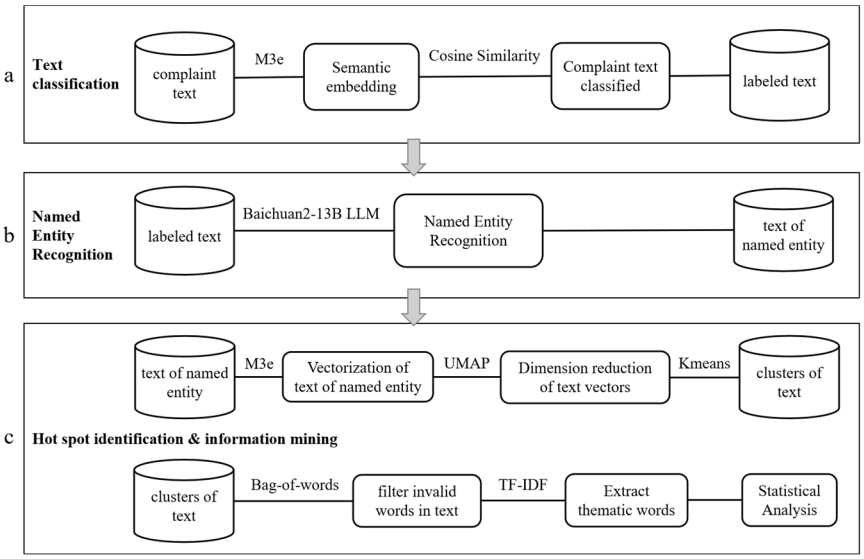

Fig. 2. Technology Roadmap

2.3.3 Hotspot identification and information mining

The process of hotspot identification and information mining is illustrated in Fig.2c. The first step involves text embedding, where we refined the original BERTopic framework by substituing the default pre-trained multilingual model (distiluse-base-multilingual-cased-v1) with the M3e model for long text embedding, leveraging the advantages of M3e as discussed earlier. The next step is dimensionality reduction by the UMAP algorithm for enhancing the information density and shortening the time required for subsequent clustering. Subsequently, the K-Means clustering algorithm is applied to identify hotspots by grouping texts into appropriate clusters. Each cluster corresponds to the same class of hotspots. Finally, drawing from the Bag-of-Words model, the texts within each cluster undergo tokenization, during which we use a set of stop words to filter invalid thematic words. And weighted extraction of thematic words representing the content of each type of hotspot complaint is conducted by TF-IDF algorithm. Based on the identified hotspot types and the thematic word prompts, we construct labels for hotspot issues, representing the prevalent concerns in each identified hotspot category.

3 Results and discussion

3.1 Classification of complaint events

According to experts in the field of social governance, complaints related to the Asian Games can be classified into five categories: "housing and economic disputes", "noise and environmental pollution", "urban planning and public facilities", "government administrative services", and "road traffic". Using the M3e-base model and cosine similarity algorithm, the complaints content was classified. Since the similarity between the embedding vectors of the complaints content and the category description follows a normal distribution, the 3σ principle was adopted to classify complaints with a maximum similarity less than 0.69 as "Others". As shown in Fig.3, it can be observed that the category "Others" accounts for a small proportion, only 3%. According to the classification results shown in Fig.3, the majority of complaint incidents are about the category of "urban planning and public facilities", followed by "government administrative services", both of which account for over 20%. The third tier category includes "road traffic", "noise and environmental pollution", and "housing and economic disputes", each of which accounts for more than 10%. Based on the category distribution of complaints, it can be preliminarily concluded that the preparation and competition period of the 19th Asian Games is a time when claims are concentrated in the categories of urban construction and government services. This may be linked to high expectation for improving cityscape.

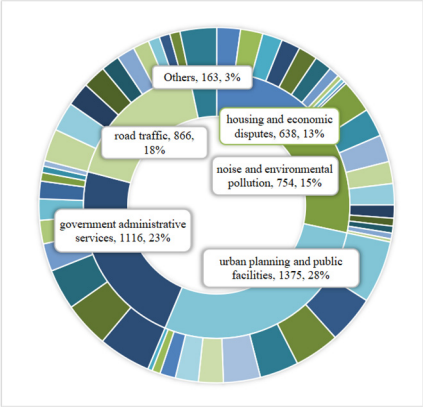

Fig. 3. Categorization distribution of complaint

3.2 Analysis results of complaint hotspots

The clustering results are measured using the "hotness index", which refers to the number of incidents in a cluster category. Higher value refers to a greater hotness of a complaint category. The top five ranking topics in each category have been retained. In this step, 25 categories of hotspots were extracted from thousands of complaints. Taking "housing and economic disputes" as an example, the top five hot topics are "hidden danger _safety_elevator_use", "community_street_trash_demolition", "capital_regulation_property_ rotten", "delivery_stoppage_extension_date", and "regulation_sale_property_department_ illegal". The overall results of the complaint type hotspot analyses are shown in Appendix 1.

3.3 Complaint expression pattern analysis

This section will analyze the expression patterns of public complaints during the international event (Asian Games) by combining the domain knowledge of social governance to further identify the features of the calculated hot topics.

The UMAP algorithm is used for dimensionality reduction with the following parameter settings: n_components=5 for the number of dimensions after reduction, n_neighbors=15 for the number of neighboring points, min_dist=0.1 for the allowed embedding density level, and metric="cosine" for calculating the cosine distance. The KMeans algorithm is used for clustering with the parameter setting of n_clusters=10 for the number of clusters. After applying the TF-IDF algorithm, the 4 most heavily weighted term is extracted as the thematic words, which are used to represent the label name of each hotspot.

The rules for adding feature labels to hot topics around the dimensions of "relevance" and "publicity" are as follows: a) If a complaint item is directly caused by the preparation or b) organization of the Asian Games and the content of the complaint has a direct impact on the organization of the Asian Games, it is considered to have strong relevance. Otherwise, it is considered to have weak relevance. If a complaint is closely related to the public interest, it is considered to have a strong public nature, and vice versa. The results of labeling the characteristics determine the placement of different hot issues in the four quadrants of Fig.4.

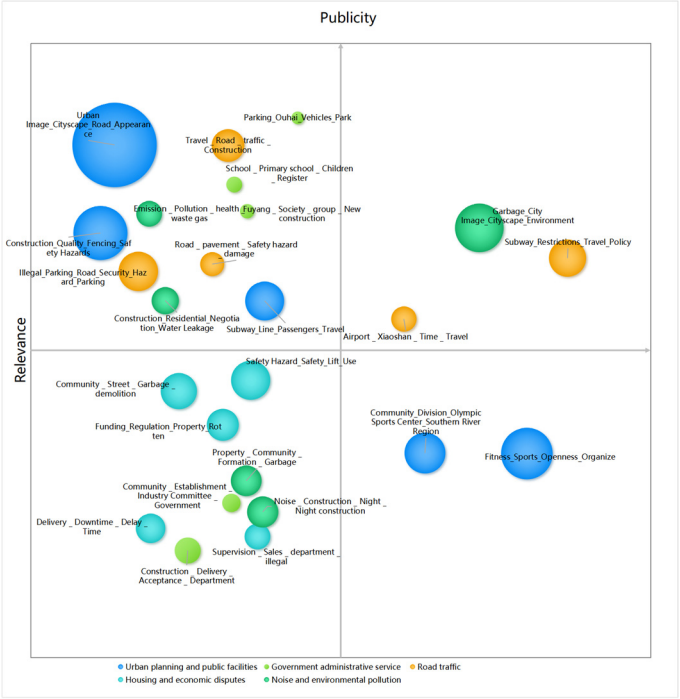

Fig. 4. The distribution of characteristics of complaint issues[1]

[1] The size of the circles in the figure represents the heat level in line with the various types of incidents.

Through in-depth analysis, we have identified common characteristics in the hotspot complaints across the four quadrants. After further condensing (as shown in Table 1), we have categorized the public's complaints related to the Asian Games into four categories: Embellishing, Threatening, Suggesting, and Complaining.

Table 1. Characteristics and Patterns of Expression of Asian Games-related Complaint Hot Spots

Hotspot	Example	Expression Characteristics	Expression Patterns
Garbage_City Image_Cityscape_Environment Metro_Restrictions_Travel_Policy ...	1.The upcoming Asian Games require the rectification of the river's odour. 2.The policy of restricting metro traffic during the competition affects travel and requires adjustment.	The public's demands on governors in terms of expectations for a good Asian Games.	Suggesting (I)
Parking_Ouhai_Vehicles_Park Illegal_Parking_Road_Security_Hazard_Parking ...	1.The parking charges at the Government Construction Ouhai Centre of Asian Games Park are causing concern. 2.Illegal parking on the roads is having a serious impact on the appearance of the Asian Games and poses security risks.	To underscore the significance of public affairs, the public employs the "Asian Games" as an embellishing tool.	Embellishing (II)
Safety Hazard_Safety_Lift_Use Funding_Regulation_Property_Rotten ...	1.The frequent failure of the Zixi Huayuancheng lift poses a threat to the safety of residents' lives and property. Additionally, it does not align with the spirit of the Asian Games Hangzhou. 2.The emergence of risks such as abandoned real estate projects and fund misappropriation during the Asian Games may potentially lead to collective petitions.	To emphasise the urgency of personal issues, the public uses the Asian Games as a tool to threaten (the government).	Threatening (III)
Community_Division_Olympic_Gangnam Fitness_Sports_Open_Organised	1.Disputes have arisen about the planning of the Olympic Community 2.The postponement of the Asian Games has affected the normal opening of stadiums. …	Public dissatisfaction due to the compromise to the Asian Games.	Complaining (IV)

Combined with the distribution of hotspot events in Fig.4, it is implicit that:

Regularity 1 - Although people mentioned the Asian Games in their complaints, most of their complaints were weakly related to the sporting event and showed more patterns of Embellishing and Threatening.

Regularity 2 - In the category of "housing and economic disputes", the term "Asian Games" is often used as a means of threatening the government. To draw the government's attention and prompt a swift resolution, complainants may link their complaint to the safety of the Asian Games or express it in a negative or harsh manner.

Regularity 3 - Issues pertaining to "Urban planning and public facilities" occupy the largest proportion in both "Embellishing" and "Complaining" types of complaints. And the expressive characteristics varies depending on whether they are related to individual interests or not.

Regularity 4: "Suggesting-type" complaints mainly come from the public who reflecting a kind of hidden problems directly affecting the Asian Games, and requesting to solve them as soon as possible, which to a certain extent reflects the social responsibility of the citizens within the hosting city of the Asian Games.

During the international sporting events, urban governance aims not only to address potential societal risks related to facilities, services, and public sentiment that may affect the smooth running of the event, but also to minimize the disruption to the daily lives of the public. Among aforementioned expression patterns associated with complaints related to the Asian Games, regularity 2 poses a potential risk and require the authorities to smooth public emotions promptly to prevent the redirection of grievances towards public events. Regularity 3 recommends that the government adjust its policies to minimize the impact of public events on people's livelihoods. According to regularity 4, the government should encourage public participation in the preparation of major events, allowing them to exercise their sense of ownership.

4 Conclusion

This study is grounded in the online complaint data from the unified platform "Min Hu Wo Wei" in Zhejiang Province. It endeavors to identify issues of public concern from complaints associated with the "Asian Games". Firstly, this study utilizes the M3e embedding and cosine similarity methods to categorize the complaints into five categories: "housing and economic disputes", "noise and environmental pollution", "urban planning and public facilities", "government administrative services", and "road traffic". Next, the BERTopic framework is used to identify the hotspots for each category. And through hotspot information mining techniques, thematic words (as the basis for hotspot naming) and representative events for the hotspots are identified as outlined in Table 2. Finally, by using a 2x2 quadrant composed of the elements "publicity" and "relevance", four expression patterns of public complaints are identified as Embellishing, Threatening, Suggesting, and Complaining. The distribution of different expression patterns is analyzed to understand the complaint regularity. This research contributes to assisting the government in accurately discerning public demands within the context of international sporting events, providing a rational foundation for the formulation reasonable social governance strategies to ensure the smooth hosting of such events.

Acknowledgments. This research is supported by the project from the National Key Research and Development Program (2022YFC3303103) of the Ministry of Science and Technology of China.

Appendix I

Table 2. Hot spot analysis and examples

Housing and economic disputes		
Name	Hot index	Example
Safety Hazard_Safety_Lift_Use	158	The frequent failure of the Zixi Huayuancheng lift poses a threat to the safety of residents' lives and property. Additionally, it does not align with the spirit of the Asian Games Hangzhou.
Community _ Street _ Garbage _ demolition	146	The placement of garbage stations has sparked controversy and the community's handling of the issue has been inadequate.
Funding_Regulation_Property_Rotten	129	The emergence of risks such as unfinished real estate projects and fund misappropriation during the Asian Games may potentially trigger collective petitions.
Delivery _ Downtime _ Delay _ Time	119	The handover of Baolong Canglongfu residential project is delayed.
Supervision _ Sales _ department _ illegal	104	The developer of Xixi Chenyue project illegally diverted sales proceeds, resulting in a prolonged project suspension.
Noise and environmental pollution		
Name	Hot index	Example
Garbage_City Image_Cityscape_Environment	196	The upcoming Asian Games require the rectification of the river's odour.
Noise _ Construction _ Night _ Night construction	126	Construction noise around Xinhe Jiayuan is disturbing to residents.
Property _ Community _ Formation _ Garbage	124	The community disregards complaints about poor neighborhood environment and the establishment of garbage stations.
Construction_Residential_Negotiation_Water Leakage	110	The ongoing construction project by Xiaoke has serious quality and noise issues, but the developer refuses to engage in negotiations.
Emission _ Pollution _ health _ waste gas	104	The industrial park in Dongqiao Road, Fuyang, Hangzhou, is emitting exhaust gases.
Urban planning and public facilities		
Name	Hot index	Example
Urban Image_Cityscape_Road_Appearance	340	Appeal for optimizing road guardrails in Hangzhou;Chaotic parking of non-motor vehicles at the entrance of Binjiang Tianjie.
Construction_Quality_Fencing_Safet	218	The northern wall of Jinhuyijingcheng is

y Hazards		significantly inclined and fractured due to the impact of elevated piling.
Fitness_Sports_Openness_Organize Community_Division_Olympic	208	The postponement of the Asian Games has affected the normal opening of stadiums.
Sports Center_Southern River Region	165	Disputes have arisen about the planning of the Olympic Community.
Subway_Line_Passengers_Travel	157	Demands for the northern extension of Metro Line 6.
Government administrative service		
Name	Hot index	Example
Construction _ Delivery _ Acceptance _ Department	105	Complaints about the negligence of relevant departments in the supervision of residential construction.
Community _ Establishment _ Industry Committee _ Government	74	Issues with property management and disputes in the election of the homeowners' committee.
School _ Primary school _ Children _ Register	64	Difficulties in transferring school for children of migrant families; The issue of education for children of talented individuals.
Fuyang _ Society _ group _ New construction	58	Destruction of forests in Fuyang for the construction of graves.
Parking_Ouhai_Vehicles_Park	51	The parking charges at the Government Construction Ouhai Centre of Asian Games Park are causing concern.
Road traffic		
Name	Hot index	Example
Illegal_Parking_Road_Security_Hazard_Parking	159	Illegal parking on the roads is having a serious impact on the appearance of the Asian Games and poses security risks.
Subway_Restrictions_Travel_Policy	150	The policy of restricting metro traffic during the competition affects travel and requires adjustment.
Travel _ Road _ traffic _ Construction	132	Road construction affects traffic.
Airport _ Xiaoshan _ Time _ Travel	102	The issue of road closures during Asian Games rehearsals affecting travel; management issues at Xiaoshan Airport.
Road _ pavement _ Safety hazard _ damage	97	The sinking and potholes on the road surface have caused safety hazards.

References

[1] Zhang, Q., Phang, C. W. & Zhang, C. Does the internet help governments contain the COVID-19 pandemic? Multi-country evidence from online human behaviour. Government Information Quarterly, vol. 39, no. 4, (2022)

[2] Chen, B., Liu, T., Guo, L. & Xie, Z. The disembedded digital economy: Social protection for new economy employment in China. Social Policy and Administration, vol. 54, no. 7, Art. no. 7 (2020)
[3] Bannykh, G. Gender Digital Inequality: Conceptualization and Practices. in Lecture Notes in Information Systems and Organisation, pp. 167–181. (2021)
[4] Zhang, Z., Lin, X. & Shan, S. Big data-assisted urban governance: An intelligent real-time monitoring and early warning system for public opinion in government hotline. Future Generation Computer Systems. vol. 144, pp. 90–104 (2023)
[5] Wang, D. & Guo, J. The Big Data Analysis and Visualization of Mass Messages under "Smart Government Affairs" Based on Text Mining. Mathematical Problems in Engineering. vol. 2022, pp. 1–18 (2022)
[6] Chen, B. & Xue, G. Research of Intelligent Analysis based on 12328 Transportation Service Supervision Hotline Data. in 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), Xiangtan, China: IEEE, pp. 1147–1152. (2023)
[7] Pu, X. et al. A Semantic-Based Short-Text Fast Clustering Method on Hotline Records in Chengdu. in 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress, Fukuoka, Japan: IEEE, pp. 516–521 (2019)
[8] Bastani, K., Namavari, H. & Shaffer, J. Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. Expert Systems with Applications. vol. 127, pp. 256–271 (2019)
[9] Li, R., Ou, R. & Wang, D. An Innovative Similar Complaint Recommendation Model Integrating Semantic and Graph Embeddings. in 2023 4th International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Guangzhou, China: IEEE, pp. 107–111 (2023)
[10] She, X., Chen, J. & Chen, G. Joint Learning With BERT-GCN and Multi-Attention for Event Text Classification and Event Assignment. IEEE Access, vol. 10, pp. 27031–27040 (2022)
[11] Si, Y., Xu, L., Peng, X. & Liu, A. Comparative Diagnosis of the Urban Noise Problem from Infrastructural and Social Sensing Approaches: A Case Study in Ningbo, China. IJERPH. vol. 19, no. 5, p. 2809 (2022)
[12] Sun, Y., Ji, M., Jin, F. & Wang, H. Public Responses to Air Pollution in Shandong Province Using the Online Complaint Data. IJGI vol. 10, no. 3, p. 126 (2021)
[13] Li, C., Huang, Y., Shen, Y. & Xu, L. Spatiotemporal patterns and mechanisms of street vending from the social sensing perspective: A comparison between law-enforcement reported and residents complain events. Cities. vol. 124, p. 103597 (2022)
[14] Rohan, T. J., Wondolowski, N. & Shelef, E. Landslide susceptibility analysis based on citizen reports. Earth Surf Processes Landf. vol. 46, no. 4, pp. 791–803 (2021)
[15] Kim, T. N., Decuir, M., Smith, K., Medus, C. & Hedberg, C. W. Use of Online Consumer Complaint Forms to Enhance Complaint-Based Surveillance for Foodborne Illness Outbreaks in Minnesota. Journal of Food Protection. vol. 86, no. 6, p. 100095 (2023)
[16] Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv (2019)
[17] 'moka-ai/m3e-base'. [Online]. Available: https://huggingface.co/moka-ai/m3e-base
[18] Yang, A. et al. Baichuan 2: Open Large-scale Language Models. arXiv (2023)
[19] Abuzayed, A. & Al-Khalifa, H. BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique. Procedia Computer Science. vol. 189, pp. 191–194 (2021)