Article

Perceiving Residents' Festival Activities Based on Social Media Data: A Case Study in Beijing, China

Bingqing Wang ¹, Bin Meng ^{1,*}, Juan Wang ¹, Siyu Chen ¹ and Jian Liu ²

¹ College of Applied Arts and Sciences, Beijing Union University, No. 197 Beitucheng West Road, Beijing 100191, China; 20201070510116@buu.edu.cn (B.W.); wangjuan@buu.edu.cn (J.W.); 20201070510119@buu.edu.cn (S.C.)

² College of Resource Environment and Tourism, Capital Normal University, No. 105 West 3rd Ring Road North, Beijing 100048, China; 2190902031@cnu.edu.cn

* Correspondence: mengbin@buu.edu.cn

Abstract: Social media data contains real-time expressed information, including text and geographical location. As a new data source for crowd behavior research in the era of big data, it can reflect some aspects of the behavior of residents. In this study, a text classification model based on the BERT and Transformers framework was constructed, which was used to classify and extract more than 210,000 residents' festival activities based on the 1.13 million Sina Weibo (Chinese "Twitter") data collected from Beijing in 2019 data. On this basis, word frequency statistics, part-of-speech analysis, topic model, sentiment analysis and other methods were used to perceive different types of festival activities and quantitatively analyze the spatial differences of different types of festivals. The results show that traditional culture significantly influences residents' festivals, reflecting residents' motivation to participate in festivals and how residents participate in festivals and express their emotions. There are apparent spatial differences among residents in participating in festival activities. The main festival activities are distributed in the central area within the Fifth Ring Road in Beijing. In contrast, expressing feelings during the festival is mainly distributed outside the Fifth Ring Road in Beijing. The research integrates natural language processing technology, topic model analysis, spatial statistical analysis, and other technologies. It can also broaden the application field of social media data, especially text data, which provides a new research paradigm for studying residents' festival activities and adds residents' perception of the festival. The research results provide a basis for the design and management of the Chinese festival system.

Keywords: social media data; festival activities; citizen perceptions; word frequency analysis; topic analysis

Citation: Wang, B.; Meng, B.; Wang, J.; Chen, S.; Liu, J. Perceiving Residents' Festival Activities Based on Social Media Data: A Case Study in Beijing, China. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 474. <https://doi.org/10.3390/ijgi10070474>

Academic Editors: Yeran Sun and Shaohua Wang

Received: 28 May 2021

Accepted: 8 July 2021

Published: 10 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Festivals are one of the representative cultures of a country, nation or region. They have multiple functions such as gathering social consensus, inheriting traditional culture, and enriching spiritual life [1]. In the process of modernization and globalization, and with economic growth, civic living conditions have gradually improved. Associated with this, the life-autonomy of city residents has increased, and options for festival activities have increased [1,2]. How to provide full access to and enrich the diverse culture and functions of traditional Chinese festivals and the revival of traditional Chinese culture are issues of current general concern to Chinese society.

In the process of globalization and modernization, conflicts and exchanges between different cultures have gradually increased. There are relatively few studies on the use of big data applied to festival cultural perception. More commonly used research frameworks to investigate folk customs and social activities are based on analysis during the actual situation, combined with theoretical analysis, and these are then used to suggest management options [3,4]. There have been many studies of the inheritance and development of

Chinese festivals from perspectives and these have provided suggested improvements. Zhang proposed that Chinese festivals are in an era of development and made suggestions on the design of Chinese festivals from the perspective of history and folklore [1]. Wang briefly reviewed the inheritance and development of Chinese traditional festivals in Hong Kong, Macau, and Taiwan [5]. Li used methods such as field investigations, questionnaire surveys, and literature studies to analyze the status of traditional Chinese festivals and to propose further developments [6]. However, it should be noted that the method of collecting relevant information based on field trips, questionnaire surveys, or interviews has high time and money costs, and is subject to questionnaires [7]. Restrictions on design, interview rules, and personal subjective factors had greatly affected the accuracy of the data, and because the temporal and spatial scale of the sample coverage was small, there was a certain risk to the reliability of the data and conclusions. Since most of the research on festivals and culture's research methods were based on field trips, questionnaire surveys, literature research, and other methods [8,9].

With the widespread adoption of mobile devices and location-based services, social media data have increasingly attracted the attention of scholars due to their large user base, rich spatiotemporal and semantic information, and low cost of access [10,11]. Meanwhile, understanding how conversational discourse on online social networks changes semantically and geographically over time will help reveal the dynamic changes of interpersonal relationships and digital traces of social events [12]. Xie and others used sign-in data of the Sina Weibo social media platform in Beijing in 2016. They used the TF-IDF (term frequency-inverse document frequency) algorithm based on geographic location information and spatial clustering to locate hot spots in Beijing in order to study social and cultural differences and crowd behaviors between different areas of Beijing [13].

Studying people's behavior is of great importance to urban planning and design and to the improvement of the living standards of residents [14,15]. Traditional methods of collecting human behavior data such as surveys are only suitable for small sample research projects. Moreover, these methods are time-consuming and costly, and the results obtained are difficult to update. In recent years, people are willing to disclose useful personal information on social media [16]. How to fully mine social media data to obtain residents' opinions of festivals has become an important topic of current research. Garay used social media (especially Twitter) to analyze the potential contribution of festivals in generating the image of festival destinations, but their research goals were more focused on the commercial value of festivals [17]. Zhou selected Sina Weibo data from 2012 to 2014 related to the five traditional festivals of the Spring Festival, the Lantern Festival, the Qingming Festival, the Dragon Boat Festival, and the Mid-Autumn Festival. People's perception of traditional Chinese festivals and regional differences in their perception of traditional festivals were investigated using word frequency analysis and LDA theme analysis [18].

Existing related research has achieved important results in research on festival activities and human perceptions [10,19]. Liu and others used social media data to study the daily activities of residents. Based on this, the proposed framework integrates textual semantic analysis, statistical method, and spatial techniques, broadens the application areas of social media data, especially text data, and provides a new paradigm for the research of residents' activities and spatiotemporal behavior [20]. However, there are relatively few studies on the analysis of residents' festival activities from two aspects, text mining and space analysis. Therefore, there is still a lot of room for research on festival activities based on social media data. We process the unpredictable, sparse, and irregular data that appears in location-based social networks, and convert this uncertain, noisy geo-tagged data into useful, well-structured high-level information [21,22] (for example, the space distributed for festival events). Minatel proposed that when using stay points to construct LBSN, it presents much more information since GPS logs convey more users' mobility information [23]. It is a very challenging task to easily explain this, to make better decisions for further festival construction. There were relatively few researches using big data from

this perspective. Therefore, there is still a lot of room for research on festival activities based on social media data.

On a small scale, such as that of a region or city, the comparison of residents' perceptions of various festivals needs further research. With the widespread adoption of mobile devices and location-based services, social media data have increasingly attracted the attention of scholars due to their large user base, rich spatiotemporal and semantic information, and low cost of access [10,11]. Meanwhile, understanding how conversational discourse on online social networks changes semantically and geographically over time will help reveal the dynamic changes of interpersonal relationships and digital traces of social events [12]. Xie and others used sign-in data for the Sina Weibo social media platform in Beijing in 2016. They used the TF-IDF (term frequency-inverse document frequency) algorithm based on geographic location information and spatial clustering to locate hot spots in Beijing in order to study social and cultural differences and crowd behaviors between different areas of Beijing [13].

Using big data analysis and text mining research methods, it is possible to examine the attitudes, activities, and preferences of people in different areas of a city, and reveal social, cultural, and functional characteristics of hot spots [24,25]. Such research methods can also be used to enhance cultural perception, to explore cultural connotations of traditional Chinese festivals in order to revive traditional Chinese festivals, and to provide suggestions and solutions to meet the requirements of the current era [26].

Using social media data from the Sina Weibo platform, based on the text and spatial temporal information, the residents' festival activities are studied from two aspects: text mining and spatial analysis. Through the integration of natural language processing technology, spatial analysis, statistical analysis, and other technical means, it provides a new research paradigm for festival culture research. This research focuses on the behavioral characteristics of Beijing residents' festival activities and their perceptions of various types of festivals. Firstly, the behaviors of festival activities are classified by extracting keywords and other information from Weibo text. The spatial patterns of various actions are then mapped. This research discussed the sensing and spatial characteristics of residents' festival activities.

The rest of this article is organized as follows. In Section 2, data collection and research methods are introduced. In Section 3, the results of sorting and categorizing the information of residents' festival activities are described, and the semantic characteristics, perceived content, and temporal and spatial patterns of residents' festival activities are analyzed. In Section 4, the advantages and disadvantages of the research methods used in this article are discussed. Finally, in Section 5, we summarize our study, draw conclusions, and propose future research directions.

2. Data and Methods

2.1. Study Area

Beijing is the capital of the People's Republic of China, a national central city, and a mega city. The Chinese Political Center, Cultural Center, International Exchange Center, and Science and Technology Innovation Center approved by the State Council. As of 2018, the city had 16 districts with a total area of 16,410 square kilometers. At the end of 2019, the permanent population was 21.536 million and the urban population was 18.65 million. The urbanization rate was 86.6%. The GDP of Beijing area was 3537.13 billion Yuan. The added value of the tertiary industry accounted for 83.5% of the regional GDP [27]. Beijing was rated as the world's first-tier city by the Globalization and World Cities Research Network (GaWC) [28]. According to the seventh Chinese national census data, among the total 21.893 million permanent residents in Beijing, the population aged 0–14 is 11.9%; the population aged 15–59 is 68.5%; and the population aged 60 and above is 19.6% [29]. Beijing is an ancient capital with a history of more than 3000 years, with rich historical and cultural heritage. The city is also a symbol and image of China, and a primary window to show China to the world. It has always attracted great attention at home and abroad.

2.2. Data

Sina Weibo is a social media platform with a large amount of social media data. According to the Sina Weibo User Development Report 2020 [30], the number of monthly active users of the software reached 511 million. Statistics from the Weibo Data Center in December 2020 show that Sina Weibo has a very high coverage rate of the city population in first-tier cities such as Beijing, Shanghai, Guangzhou, and Shenzhen [30]. The Sina Weibo data contains a considerable amount of various geographic information. Through the Sina Weibo software and web crawler tools, we obtained Sina Weibo data from Beijing for the year 2019 and captured the content of Weibo posts in a targeted manner. The data included Weibo ID, latitude and longitude, time, mobile terminal, region, text content, and other information. In total, more than 1.13 million individual pieces of data were obtained as the data source for this study (Figure 1).

Figure 1. The location of Weibo data on festivals in Beijing.

2.3. Methods

The framework of residents' festival activities research as follows (Figure 2).

Figure 2. Framework of residents' festival activities research based on social media data.

2.3.1. Semantic-Based Weibo Classification and Extraction

A text classification model based on BERT and Transformers framework were built in this research. The BERT model was one kind of language encoder, released by Google in

2018, able to translate the input sentences or paragraphs into corresponding semantic features, which has performed amazingly well and become an important recent advancement in NLP [31]. In this research, we used the Simple Transformers library [32], which is based on the Transformers library by HuggingFace [33], to build our model. The model can be quickly trained and evaluated.

First, on the basis of cleaning 1,136,125 Weibo posts (removing tags, attaching emails, forwarding links, expressions, videos, sharing pictures, and other information unrelated to the content of the text), the pre-trained model of BERT-base-Chinese was initialized to do binary classification. Secondly, 7000 randomly selected posts as training samples were used to train the model. For each post, if it was related to the residents' festival activities, it was marked as 1, otherwise it was marked as 0. Machine learning and the original BERT model were then used to verify the classification accuracy. By adjusting the corresponding parameters and the number of iterations several times under the experiment, a trained text multi-classification model was obtained (the model accuracy reached 97%). Third, based on the derived classifier, all Weibo entries were input into BERT to classify Weibo with residents. After classification and extraction, there were 213,649 social media posts related to the festival.

2.3.2. Word Frequency Statistics

The word frequency statistics based on the TF-IDF algorithm are to evaluate the importance of a word to a text. If a certain word or phrase appears frequently in an article, and appears low in the document collection, this word or phrase is considered as having good ability to distinguish categories [34].

Specifically, first, the Weibo data was segmented based on Jieba. The purpose is to split the words in the text and transform the text into multiple words in order. Word segmentation was equivalent to feature extraction, and the extracted words were called feature words. After obtaining the characteristic words, this research then used the custom dictionary and stop word database to filter out some prepositions and symbols because the text was more complex and the word content was large. Finally, the characteristic words were selected, which played a major role in text classification and topic analysis, and ranked them in order of importance.

2.3.3. Topic Model LDA

Latent Dirichlet allocation (LDA) is one of typical "bag of words" model [34] and has a wide range of applications [35]. It is a standard topic model that can work with social media data where there is a problem of short text and large sparseness [36]. Its basic idea is that the text is randomly mixed and generated from implicit topics, and each topic corresponds to a specific feature word distribution [37].

This study constructed a three-layer Bayesian structure of "text-topic-word" based on social media data. The topic of each text in the text set is given in the form of a probability distribution, so as to classify topics according to the topic distribution. This research attempted to create a list of topics through the results to examine the spatial characteristics of Beijing residents' festival activities and to visualize the results [38].

2.3.4. Spatial Analysis

Spatial analysis is a widely used analysis method in geography [39]. In this study, the main focus is on the spatial distribution of data. Related methods include density analysis, spatial interpolation analysis, spatial visualization, and measurement of geographic distribution [40]. Festival-related Weibo content was displayed in space through topic clustering and kernel density analysis was performed to observe the hot spots in the space.

3. Results

3.1. Festive Event Word Frequency Statistics

Festivals with more than 10,000 Weibo posts were National Day, Mid-Autumn Festival, New Year's Day, Christmas Day, Lantern Festival, and Christmas Eve (Table 1). As 2019 was the 70th anniversary of the founding of the People's Republic of China, most Weibo posts were related to the National Day. The status of the family in the Chinese people's concept of festival is culturally important, and hence the Mid-Autumn Festival with the theme of a family reunion was the second largest festival-related Weibo content in 2019.

Table 1. Festivals with more than 10,000 Weibo posts in 2019 and their dates.

Festival	Date	Number of Related Weibo Posts	Number of Weibo Posts on the Day
National Day	1 October 2019	29,157	69,820
Mid-Autumn Day	13 September 2019	21,807	55,440
New Year's Day	1 January 2019	15,211	57,145
Christmas Day	25 December 2019	13,089	42,766
Lantern Festival	19 February 2019	11,172	34,121
Christmas Eve	24 December 2019	10,720	40,565

All of the 213,649 festival-related Weibo posts from Beijing in 2019 were sorted by word frequency statistics (Figure 3). As 2019 was the 70th anniversary of the founding of the People's Republic of China, the frequency of words related to the National Day, such as “motherland”, “happy birthday”, “70”, was high. The frequency of entries related to the Mid-Autumn Festival was also high. In Chinese festival activities, eating food was clearly an essential behavior and a principal way people participated in festivals.

Figure 3. Word frequency map.

Based on all festival-related Weibo content in 2019, the main content of residents' perception of festivals and the main ways of participating in festivals were reflected in word cloud diagrams (Figure 4). High-frequency words corresponded to festivals with a large number of Weibo posts in 2019. For example, words such as “motherland”, “China”, and “happy birthday” were also reflected in word cloud maps for National Day, Mid-Autumn Festival, New Year, and other related words. Words such as “eat” and “delicious” reflected the main ways that residents participate in festivals.

Figure 4. Word cloud illustration from festival-related Weibo posts.

The festivals were divided into three categories: traditional festivals, foreign festivals, and modern festivals, sorted according to the number of related posts from most to least, and the proportion of the number of posts of various types of festivals in the total data were calculated. The results were shown in Table 2.

Beijing residents posted the largest number of Weibo posts related to traditional festivals, accounting for 40.46%. Among the traditional festivals, the Mid-Autumn Festival, with the theme of a family reunion, was the most frequently mentioned. However, the number of Weibo posts related to the Spring Festival was relatively small. This was due to the fact that the time span of the Spring Festival was long. Only the Weibo data on the day of the holiday was extracted here, so there was a deviation in the number of Weibo posts. In addition, Weibo users tend to be young, and hence the Weibo post data may not reflect the feelings of middle-aged and elderly people.

Table 2. The number and proportion of festival-related Weibo posts made in Beijing in 2019, split by festival type and individual festival.

Festival Type	Festival Name	Number of Weibo Posts	Percentage of Festival Ratio	Weibo Posts Proportion of Festival Types
Traditional festival	Mid-Autumn Day	21,807	25.23%	40.46%
	The Lantern Festival	11,172	12.93%	
	The Dragon Boat Festival	8861	10.25%	
	The Spring Festival	8734	10.10%	
	Chinese New Year’s Eve	8247	9.54%	
	Double-ninth Day	7252	8.39%	
	Ching Ming Festival	6301	7.29%	
	The Laba Festival	5716	6.61%	
	Double Seventh Festival	4415	5.11%	
	Chinese Yuan Festival	3932	4.55%	
	Subtotal	86,437	100%	
Foreign festival	Christmas Day	13,089	30.03%	20.40%
	Christmas Eve	10,720	24.59%	
	Valentine’s Day	6091	13.97%	
	Halloween	4808	11.03%	
	Easter Day	3945	9.05%	
	April Fools’ Day	2768	6.35%	
	Thanksgiving	2169	4.98%	
		Subtotal	43,590	

Table 2. Cont.

Festival Type	Festival Name	Number of Weibo Posts	Percentage of Weibo Posts	
			Festival Ratio	Proportion of Festival Types
Modern festival	The National Day	29,157	34.87%	39.14%
	New Year's Day	15,211	18.19%	
	International Children's Day	6981	8.35%	
	Mother's Day	5822	6.96%	
	International Labour Day	5599	6.70%	
	Chinese Youth Day	4482	5.36%	
	Father's Day	3656	4.37%	
	The Army's Day	3628	4.34%	
	Teachers' Day	3294	3.94%	
	The Party's Birthday	2750	3.29%	
	Arbor Day	2039	2.44%	
	International Working Women's Day	1003	1.20%	
	Subtotal	83,622	100%	
Total	29	213,649	100%	100%

Traditional festivals are closely related to Chinese history and culture. In order to explore the degree of attention to traditional culture in Weibo, it is necessary to analyze some relatively low-frequency words in the characteristic words (Table 3). Residents' festival activities are greatly influenced by traditional culture. This is not only reflected in clothing and locations, such as "Hanfu" and "Confucian Temple". In traditional festivals, the influence of traditional culture is more obvious. "Will live long as he can!", "From far away you share this moment with me." and other phrases corresponding to the Mid-Autumn Festival appear more frequently.

Table 3. Traditional culture-related word frequencies.

Words Related to Traditional Culture	Frequency
He who does not reach the Great Wall is not a true man. (不到长城非好汉)	255
blooming flowers and full moon (花好月圆)	155
On festive occasions more than ever one thinks of one's dear ones far away. (每逢佳节倍思亲)	144
Will live long as he can! (但愿人长久)	141
A fall of seasonable snow gives promise of a fruitful year. (瑞雪兆丰年)	72
Hanfu (汉服)	64
From far away you share this moment with me. (天涯共此时)	57
Confucian Temple (孔庙)	53

Foreign festivals accounted for 20.40% of the Weibo data on the day of the festival, showing that traditional festivals still dominate residents' perception of festivals. Besides Christmas and Christmas Eve being key points of residents' sense of foreign festivals, foreign festivals do not occupy the central position of residents' sense of festivals. For modern festivals, the number of posts related to National Day, where residents expressed their patriotic feelings, accounted for about one-third of the total number of posts.

We also found that some festival activities, especially some foreign festivals, have a certain connection with religion (Table 4). In the published text information, not only the names of religious beliefs are clearly mentioned, but the names of religious places appear relatively frequently on the day of the festival.

Table 4. Religious-related word frequency statistics.

Word	Frequency
church	255
Buddha	155
Huguo Temple	144
Buddhism	141
Qingliang Temple	72
Fayuan Temple	64
Taoism	57
Hongluo Temple	53

3.2. Semantic Sensing of Festival Activities

Figure 5 shows the internal proportions of various types of different types of festival data, and a longitudinal comparison of the same type of data. You can find the same type of festival data, the proportions of different parts of speech and types. Especially in traditional festivals, verbs make up the largest proportion of words, which is significantly different from other types of festivals. Figure 6 is a horizontal comparison of different types of festival data of the same parts of speech. Modern festivals have the most holiday features in nouns, and traditional festivals in have “Eating” as the most common verb.

Figure 5. Percentage of part-of-speech results from festival-related Weibo posts, split by festival type. Blue shading—nouns; orange shading—verbs; green shading—adjectives.

Figure 6. Part-of-speech result statistics from festival-related Weibo posts, split by festival type.

Nouns reflected residents’ perception of festivals, especially the representative symbols and elements of festivals, for example, the nouns “moon cake”, “zongzi”, and “tangyuan”, as these traditional Chinese foods were used in relation to the traditional festivals, i.e., Mid-Autumn Festival, Dragon Boat Festival, and Lantern Festival, respectively. Words such as “Santa Claus”, “Christmas gift”, and “apple” were used, related to foreign festivals, i.e., Christmas and Christmas Eve. For modern festivals, words like “mother country” and “China”, related to National Day, were used frequently.

Regardless of the type of festival, the word “Forbidden City” appeared frequently. This indicates that the local attraction of the Forbidden City has become an indispensable

part of festivals in the perception of Beijing residents, providing an emotional support and cultural symbol. Finally, the proportion of Weibo terms of each type of festival showed that the proportion of traditional festivals was the largest, as high as 59%, which showed that residents had the most abundant perception of traditional festivals.

All high-frequency words were divided into four categories according to part of speech and semantic content. For example, such as “eat”, “drink” etc. in the group verb. In order to better summarize such activities, we named these words “eating”. Activities that can also be carried out in daily life such as “check in” and “walking around” are named “leisure activities”. Because of the limited space of other word classifications, there is not much explanation. Verbs reflect the main behaviors of residents participating in festivals. From the frequency of words, the behaviors of Beijing residents participating in festivals appeared relatively uniform across festival types (Figure 6). For example, words such as “eat” and “check in” indicate that the main behaviors of residents participating in festivals were associated with dining. It would appear that online celebrity shops’ “check in” has become an important way for Beijing residents to participate in festivals.

Adjectives mainly reflect the emotional expression of residents towards festivals, and different types of festivals corresponded to different emotional expressions. “Ching Ming” in traditional festivals corresponded to the Ching Ming Festival. Words such as “peaceful”, “smooth”, and “consummate” were cultural manifestations of traditional festivals. The word “peaceful” in foreign festivals appeared most frequently, which corresponded to people’s wish for peace on Christmas Eve. High-frequency adjectives used for modern festivals reflected the concentration of residents on the National Day, expressing pride in the motherland and giving positive comments on the status quo of the motherland, with adjectives such as “safe”, “strong”, and “prosperity”.

3.3. Spatial Distribution Characteristics of Festival Activities

Figure 7 shows the kernel density distribution map of Beijing residents’ Weibo posts related to festivals in 2019, overall and by festival type. The density distribution of traditional festivals was not much different from that of modern festivals, although the central density of residents’ postings related to traditional festivals was denser than that of modern festivals. The density of foreign festivals appeared much lower than either traditional or modern festivals, but there appeared to be many areas with no posts, suggesting that traditional festivals still occupy the main position of Chinese residents’ holiday behavior and culture. This contradicts the perception that traditional festivals are being significantly impacted and influenced by foreign festivals.

Figure 7. Kernel density map of Beijing residents' festival activities.

3.4. Theme Sensing of Festival Activities

Among the 29 festivals in 2019, the LDA theme model divided the festival-related posts into three types: the emotional expression of the posts; the specific behavior of residents; and the representative culture of the related festival. Residents' festival activities were roughly divided into two categories: eating food with relatives and friends and going to various restaurants to check-in; going to multiple tourist attractions and festival activities. The LDA model analysis was applied to the three festival types; modern, traditional, and foreign, and results were imported into ArcGIS for thematic spatial analysis.

The 5 topics, each topic was more evenly distributed in space, but topic 2 was most distributed in space (Figure 8). Comparing Table 5, the high-frequency words of topic 2 mainly correspond to the Mid-Autumn Festival and the Spring Festival, such as "moon cake", "reunion", "year of pig", and "good luck".

Figure 8. LDA spatial distribution map of traditional festivals.

Table 5. Traditional festival LDA theme high-frequency words.

Theme	High Frequency Words
1	Forbidden City, Ankang, Delicious, Lantern Festival, Friends, Moon, Hope, Taste, Peace, Happy New Year
2	Mooncake, National Day, Reunion, Blessing, Family, Holiday, Year of the Pig, Good Luck
3	New Year, clocking in, going home, rice dumplings, Tiananmen Square
4	2019, Tangyuan, New Year's Day, Laba Congee, Smooth
5	Happy new year, fifteenth lunar month

The theme space distribution of foreign festivals was not as wide as that of traditional festivals, but there are obvious spatial differences in the theme space distribution (Figure 9). Theme 1 is mainly distributed in the area outside the Fifth Ring Road in Beijing, and theme 4 is mainly distributed in the area inside the Fifth Ring Road. According to the topic high-frequency words in Table 6, topic 1 was mainly associated with residents' emotional perception and expression of the festival, with as words such as "happiness", "hope", and "peace". Theme 4 was mainly related to specific behaviors of residents participating in festivals, such as "Christmas gifts" and "apples", which means that residents participating in Christmas mainly give gifts and apples to express their care for relatives and friends.

Figure 9. LDA spatial distribution map of foreign festivals.

Table 6. Foreign festival LDA theme high-frequency words.

Theme	High Frequency Words
1	Merry Christmas, peace, hope, 2019, snowing, receiving, joy, good night, friends, merrychristmas
2	Gifts, Santa Claus, Surprise, Christmas Tree, Peace and Peace, Hot Pot, World, Restaurant
3	Forbidden City, delicious, taste, park, food
4	Check in, Christmas gifts, apples, Forbidden City

Theme 2 and Theme 3 for modern festivals also showed significant spatial differences (Figure 10). Combined with the high-frequency words in Table 7, high-frequency words in theme 2 included “Happy New Year”, “Military Parade”, “Hope”, “Fireworks”, “Tiananmen Square”, and other words, some of which expressed the best wishes of residents during the festival. The other part mainly described the representative symbols and constituent elements of festivals, especially National Day. The high-frequency words of theme 3, such as “delicious”, “check in”, and “taste” were associated with food and eating.

Figure 10. LDA spatial distribution map of modern festivals.

Table 7. Modern festival LDA theme high-frequency words.

Theme	High Frequency Words
1	Happy birthday, 2019, 70, prosperity, blessing, anniversary, I love you, birthday, forever, friend, wish, birthday
2	Happy New Year, National Day, military parade, hope, Tiananmen Square, Forbidden City, fireworks
3	Delicious, clock in, first day, new year, 2018, people, taste
4	Long live, prosperous age, June 1st, Guotai Minan

Combined with the differences in the spatial distribution of the theme of foreign festivals, it could be concluded that the main way residents participated in festivals had a certain relationship with the perfection of infrastructure. In the specific festival behaviors, residents living in the central city of Beijing can participate in various festival activities, so most of the content on Weibo reflects specific festival behaviors. Residents living in the suburbs of Beijing may have been restricted due to access to such infrastructure. Therefore, people expressed more wishes on the content of Weibo, with regard to the festival or the cultural concept of the festival itself.

4. Discussion

Most current researches on festivals and culture are conducted through surveys and field trips, and seldom uses big data to analyze related issues. Therefore, many scholars have realized the urgency of using social media data to carry out research on festival activities [4]. For example, Zhou's research mainly uses word frequency statistics and LDA theme models to identify residents' perceptions of traditional festivals and regional differences [18]. According to their research results, LDA topic classification is obviously a powerful method for analyzing social media data, text mining, and revealing the spatiotemporal characteristics of related activities. A study by Liu [41] studied the

emotional characteristics of Chinese tourists to Australia based on big data text analysis and part-of-speech tagging. These methods all extend the textual analysis of festival activities. However, the above research lacks comprehensive mining of the rich semantic information and spatiotemporal information in social media data. Therefore, this research uses NLP technology to identify festival-related Weibo posts, and combines word frequency statistics, text labeling, LDA theme models, and GIS spatial analysis methods to analyze residents' perception characteristics of festivals and activities.

Judith Mair and Karin Weber [3] pointed out that many studies in the field of festival analysis had adopted a case study approach. Therefore, the research on special festivals is relatively sufficient, but the comprehensive comparative study of many festivals is lacking. This could be said to limit the scope and scale of our understanding of festival. Therefore, by expanding the scope of research on different types of festivals, we hope to improve the understanding the residents' perception of different festivals. Through the comparison of different types of festivals, the research found that Weibo texts reflect that residents pay more attention to different festivals. Traditional festivals still receive more widespread attention; from thematic analysis, it can be found that there are common characteristics between different types of festivals. For example, attention to leisure activities and food is very prominent; it is also universal to express greetings to family and friends through festivals. However, it can also be found that among different types of festivals, traditional festivals are more closely related to history and culture, while modern festivals are more closely related to leisure and consumption. Western festivals have been more connected with consumption and entertainment while retaining some religious imprints. Such a comprehensive study is of great significance for in-depth understanding of the connotations of festivals and social and economic development.

On the spatial scale, this study found an interesting phenomenon in the spatial pattern of residents' festival activities in a giant city. Although the gathering areas for different types of festivals are concentrated in densely populated urban centers, the activities of traditional festivals and modern festivals' distribution ranges are significantly larger than that of foreign festivals in the West. We believe that the way residents participate in festivals is related to the degree of infrastructure, especially the number of entertainment facilities such as catering and services. At the same time, the regional difference of festival activities within the city also proves the imbalance in the urban structure of Beijing, that is, the northern part of the city is more distributed than the southern area (Figure 7) [42]. Just as Wilson [4] emphasized the important role of festivals to local communities. By increasing festival-related facilities in underdeveloped urban areas, it is also possible to promote the balanced development of the city. However, this research also poses a new challenge, that is, the difference between the east and the west of the city is also more obvious. This part of the reason needs to be studied in depth.

The results of this research show that we can understand the residents' perception of festivals by using social media big data. However, according to the 2020 Weibo User Development Report, Weibo users are predominantly people aged 20–30, and account for close to 80% of users [23]. Therefore, social media data is more of a relatively young group, and the data has problems with sample bias and representativeness. In order to solve this problem, in further research, traditional questionnaire surveys and other methods can be used to supplement the research samples by combining multiple sources of data to compensate for the problem of social media data sample deviation.

5. Conclusions

This study uses social media data to study residents' perceptions of festivals and the spatial characteristics of activities. By using a text classification model based on BERT and Transformers framework, we analyzed Weibo social media data related to festivals in Beijing in 2019. We obtained Beijing residents' perceptions of festivals and the ways they participated in festivals, and explored the spatial differences of residents' participation in festival activities.

Using word frequency statistics, part-of-speech analysis, and LDA topic model analysis, we analyzed Weibo social media data related to festivals in Beijing in 2019. We obtained Beijing residents' perceptions of festivals and the ways they participated in festivals, and explored the spatial differences of residents' participation in festival activities.

Traditional culture had a huge influence on festivals, which is not only reflected in residents' motivation to participate in festivals, but also in the ways they participated in festivals and the feelings they expressed. Traditional festivals occupied the central position of residents' perception of festivals. This was different from current concerns that traditional festivals are being greatly affected and impacted by foreign festivals. The feelings of family and motherland occupied a central position in modern festivals. This was clearly manifested in word frequency and topic spatial distribution. For traditional festivals, residents expressed their feelings through ancient poems from traditional Chinese culture. For example, for traditional festivals, frequently use words such as "On festive occasions more than ever one thinks of one's dear ones far away. (每逢佳节倍思亲)", "Will live long as he can! (但愿人长久)", and other verses were not used in relation to other types of festivals. The way residents participate in festivals is related to the degree of infrastructure, especially the number of entertainment facilities such as catering and services. Most of the Weibo posters from inner-city areas expressed specific festival-related behaviors showing that they were directly participating in the festival activities, while posters in the outer city area often expressed holiday wishes. Additionally, some of the residents' festival activities were related to religious beliefs, reflecting the cultural traditions and connotations behind the festivals in different types of festivals.

Through the analysis of the spatial distribution pattern of festival-related microblogs, it can be found that the temporal and spatial information of social media data can help understand the characteristics of urban spatial structure. Residents' festival activities are concentrated in densely populated and economically developed urban centers. The regional differences between the north and the south festival activities within the city are also in line with the characteristics of Beijing's urban spatial structure. However, this study found that the difference between the eastern and western parts of the city is also very obvious. This discovery presents a new challenge. The reasons for the differences between the east and west spaces of residents' activities need to be studied in depth.

This study uses social media data to study residents' perceptions of festivals and the spatial characteristics of activities. Combining natural language processing technology, statistical analysis, part-of-speech tagging, topic analysis, and spatial analysis, provides a new paradigm for the research in the field of festivals. However, the LDA topic model has certain shortcomings in processing sparse social media data. This requires subsequent advances in data processing technology. There is a problem of sample bias in social media data, which cannot reflect the situation of middle-aged and elderly people who use fewer social media well. In the follow-up research, traditional questionnaire survey methods can be used to supplement the samples with multi-source data. The spatial differences of residents' festival activities found in this study can only be described from a qualitative perspective at present. In the future, we hope that further studies can explain the reasons for the spatial differences from a quantitative perspective.

Author Contributions: Conceptualization, methodology, data curation, writing—review and editing, Bingqing Wang, Bin Meng and Juan Wang; investigation, software, visualization, writing—original and draft preparation, Bingqing Wang, Siyu Chen and Jian Liu; funding acquisition and project administration, Bin Meng and Juan Wang. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Grant Nos. 2017YFB0503605), National Natural Science Foundation of China (Grant Nos. 41671165) and the Academic Research Projects of Beijing Union University (Grant Nos. ZK40202001).

Data Availability Statement: The data is available from the authors upon reasonable request.

Acknowledgments: We would like to thank the anonymous reviewers for their insightful comments and substantial help on improving this article. We also thank Dongsheng Zhan for providing the valuable data and technical support.

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

References

1. Zhang, B. Construction of Chinese festivals in the era of construction. *Folk. Stud.* **2015**, *1*, 62–73.
2. Tai, X.C. Analysis of the Inheritance Status of Chinese Traditional Festivals and Research on Development Countermeasures. *Art Sci. Technol.* **2019**, *32*, 105–106.
3. Mair, J.; Weber, K. Event and festival research: A review and research directions. *Int. J. Event Festiv. Manag.* **2019**, *10*, 209–216. [\[CrossRef\]](#)
4. Wilson, J.; Arshed, N.; Shaw, E.; Pret, T. Expanding the Domain of Festival Research: A Review and Research Agenda. *Int. J. Manag. Rev.* **2017**, *19*, 195–213. [\[CrossRef\]](#)
5. Wang, X.W. Inheritance and development of traditional Chinese festivals in Hong Kong, Macao and Taiwan. *Cult. Herit. Bimon.* **2013**, *2*, 23–30.
6. Research Group of “Promoting Festival Culture”. Status in Quo and Development Countermeasure of Inheriting Traditional Chinese Festival. *Hundred Sch. Arts* **2012**, *28*, 1–4.
7. Wong, K.; Domroes, M. Users’ perception of Kowloon Park, Hong Kong: Visiting patterns and scenic aspects. *Chin. Geogr. Sci.* **2004**, *14*, 269–275. [\[CrossRef\]](#)
8. Schwanen, T.; Kwan, M.P. The Internet, mobile phone and space-time constraints. *Geoforum* **2008**, *39*, 1362–1377. [\[CrossRef\]](#)
9. Batty, M.; Axhausen, K.W.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; Ouzounis, G.; Portugali, Y. Smart cities of the future. *Eur. Phys. J. Spec. Top. Eur.* **2012**, *214*, 481–518. [\[CrossRef\]](#)
10. Zhang, F.; Zhou, B.; Liu, L.; Liu, Y.; Fung, H.H.; Lin, H.; Ratti, C. Measuring human perceptions of a large-scale urban region using machine learning. *Landsc. Urban Plan.* **2018**, *180*, 148–160. [\[CrossRef\]](#)
11. Liu, Y.; Yuan, Y.H.; Zhang, F. Mining urban perceptions from social media data. *J. Spat. Int. Sci.* **2020**, *20*, 51–55.
12. Koylu, C. Modeling and visualizing semantic and spatio-temporal evolution of topics in interpersonal communication on Twitter. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 805–832. [\[CrossRef\]](#)
13. Xie, Y.J.; Peng, X.; Huang, Z. Image Percept. Beijing’s Reg. Hotspots Based Microblog Data. *Prog. Geogr.* **2017**, *36*, 1099–1110.
14. Kestens, Y.; Lebel, A.; Daniel, M.; Thériault, M.; Pampalon, R. Using experienced activity spaces to measure foodscape exposure. *Health Place* **2010**, *16*, 1094–1103. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Vallée, J.; Cadot, E.; Roustit, C.; Parizot, I.; Chauvin, P. The role of daily mobility in mental health inequalities: The interactive influence of activity space and neighbourhood of residence on depression. *Soc. Sci. Med.* **2011**, *73*, 1133–1144. [\[CrossRef\]](#)
16. Marti, P.; Serrano-Estrada, L.; Nolasco-Cirugeda, A. Social Media data: Challenges, opportunities and limitations in urban studies. *Comput. Environ. Urban Syst.* **2019**, *74*, 161–174. [\[CrossRef\]](#)
17. Garay, L.; Morales, S. Understanding the creation of destination images through a festival’s Twitter conversation. *Int. J. Event Festiv. Manag.* **2017**, *8*, 39–54. [\[CrossRef\]](#)
18. Zhou, J.Y.; Wang, J.R.; Zhang, J.Q. Perception and regional differences of Chinese traditional festivals by Weibo users. *J. Geo-Inf. Sci.* **2019**, *21*, 77–85. [\[CrossRef\]](#)
19. Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social sensing: A new approach to understanding our socio-economic environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [\[CrossRef\]](#)
20. Liu, J.; Meng, B.; Wang, J.; Chen, S.; Tian, B.; Zhi, G. Exploring the Spatiotemporal Patterns of Residents’ Daily Activities Using Text-Based Social Media Data: A Case Study of Beijing, China. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 389. [\[CrossRef\]](#)
21. Carmela, C. Next: A framework for next-place prediction on location based social networks. *Knowl. Based Syst.* **2020**, *204*, 106205.
22. Hasan, M.; Orgun, M.A.; Schwitter, R. A survey on real-time event detection from the Twitter data stream. *J. Inf. Sci.* **2018**, *44*, 443–463. [\[CrossRef\]](#)
23. Minatel, D.; Ferreira, V.; Lopes, A.D.A. Local-entity resolution for building location-based social networks by using stay points. *Theor. Comput. Sci.* **2021**, *851*, 62–76. [\[CrossRef\]](#)
24. Hssan, S.; Zhan, X.Y.; Ukkusuri, S.V. Understanding Urban Human Activity and Mobility Patterns Using Large-scale Location-based Data from Online Social Media. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, Chicago, IL, USA, 11 August 2013.
25. Mayer, S.V.; Cukier, K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Reprint ed.; Houghton Mifflin Harcourt: Boston, MA, USA, 2013.
26. Liu, Y. Rethinking some basic issues of human geography from the perspective of social perception. *Acta Geogr. Sin.* **2016**, *71*, 564–575.
27. 2020 Beijing Statistical Yearbook. Available online: <http://nj.tjj.beijing.gov.cn/nj/main/2020-tjnj/zk/indexch.htm> (accessed on 25 May 2021).
28. The World According to GaWC 2018. Available online: <https://www.lboro.ac.uk/gawc/world2018t.html> (accessed on 25 May 2021).
29. Communique of the Seventh National Census of Beijing Municipality (No. 3). Available online: http://www.beijing.gov.cn/gongkai/shuju/sjd/202105/t20210519_2392888.html (accessed on 28 June 2021).

-
30. Weibo 2020 User Development Report. Available online: <https://weibo.com/ttarticle/p/show?id=2309404613871951282183> (accessed on 19 May 2021).
 31. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
 32. Available online: <https://github.com/ThilinaRajapakse/simpletransformers> (accessed on 5 July 2021).
 33. Wolf, T.; Chaumond, J.; Debut, L.; Sanh, V.; Delangue, C.; Moi, A.; Cistac, P.; Funtowicz, M.; Davison, J.; Shleifer, S.; et al. Trans-formers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Online, 5–10 July 2020; pp. 38–45.
 34. Li, R.; Zhang, W.B. Application of Data Mining Technology Based on TF-IDF Algorithm and LDA Topic Model in Power Customer Complaint Text. *Tech. Autom. Appl.* **2018**, *37*, 46–50.
 35. Gao, T.T.; Liu, W.Z.; Meng, B.; Huang, S.; Chen, S.Y. A Perception Study of the Cultural Resource-intensive Areas of the Model Based on the Theme—A Case Study of Mentougou District of Beijing. *J. Beijing Union Univ.* **2019**, *33*, 45–55.
 36. Wang, P.; Gao, C.; Chen, X.M. Research on LDA Model Based on Text Clustering. *Inf. Sci.* **2015**, *33*, 63–68.
 37. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
 38. Bahrehdar, A.R.; Purves, R.S. Description and characterization of place properties using topic modeling on georeferenced tags. *Geo-Spat. Inf. Sci.* **2018**, *21*, 173–184. [[CrossRef](#)]
 39. Zhai, J.; Jin, X.C. GIS Spatial Analysis Method in Urban Planning. *Urban Plan.* **2014**, *38*, 130–135.
 40. Fang, Y.; Yan, W. Tracking urban geo-topics based on dynamic topic model. *Comput. Environ. Urban Syst.* **2020**, *79*, 101419. [[CrossRef](#)]
 41. Liu, Y.; Huang, K.X.; Bao, J.G.; Chen, K.Q. Listen to the Voices from Home: An Analysis of Chinese Tourists’ Sentiments regarding Australian Destinations. *Tour. Manag.* **2019**, *71*, 337–347. [[CrossRef](#)]
 42. Sun, Z.; Shi, P. The Regional Difference Analysis of Urban Development in Beijing. *Urban. Dev. Stud.* **2012**, *19*, 56–59.