

Article

Analysis of Spatial Interaction between Different Food Cultures in South and North of China

——Practices from people's daily life

Haiping Zhang^{1,2,3}, Xingxing Zhou^{1,2,3,*}, Yi Huang^{1,2,3}

- 1 Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023; China; gissuifeng@163.com (H.Z.); zhouxg@163.com (X.Z.)
- 2 State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China;
- 3 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China;

Abstract: As an important research content in cultural geography, the exploration and analysis of the laws of regional cultural differences has great significance for the discovery of distinctive cultures, protection of regional cultures and in-depth understanding of cultural differences. In recent years, with the "spatial turn" of sociology, scholars are paying more and more attention to the implicit spatial information in social media data and the various social phenomena and laws they reflect. One important aspect is to grasp the social cultural phenomena and its spatial distribution characteristics through the text. Using machine learning methods such as the popular natural language processing (NLP), this paper can not only extract hotspot cultural elements from text data but also accurately detect the spatial interaction pattern of some specific cultures and the characteristics of emotions towards non-native cultures. Taking the 6,128 answers to the question "what are the differences between South and North China that you never know" on the Zhihu Q&A Platform as an example, with the help of NLP, this paper has explored the cultural differences between South and North China in people's mind. This paper probes into people's feeling and cognition of the cultural differences between South and North China from three aspects, including spatial interaction patterns of hotspot cultural elements, components of hotspot culture and emotional characteristics under the influence of cultural differences between North and South. The study reveals that 1) people from North and South China have great differences in recognizing each other's culture. 2) Food culture is the most popular among many cultural differences. 3) People tend to show negative attitude towards the food cultures different from their own. All these findings shed light upon the understanding of regional cultural differences and addressing cultural conflicts. In addition, this paper also provides an effective solution to the study from a macro perspective, which have been difficult for new cultural geography.

Keywords: Cultural differences, spatial interaction patterns, emotion analysis, Zhihu topic data, cultural geography.

1. Introduction

With the further development and changes of society, many new cultural phenomena are emerging. The study of social culture from the perspective of geography is receiving more and more attention from geographers. The development of cultural geography, as an important branch of human geography, has undergone a gradual process. Looking back at its development, some scholars believe it has undergone the transformation from conventional cultural geography to new cultural geography, which shall be the mainstream for cultural geography study [1,2]. Others hold that "new" or "old" are but two different study ideas [3]. This paper prefers the latter opinion. Meanwhile, the paper also believes that the bottom-up cultural view and the cultural meaning of daily lives and social practice stressed by new cultural geography are of great importance to the study of cultural geography [4,5]. In the past, studies from the perspective of new cultural geography mostly focus on

the analysis of cultural phenomena at the micro level. It seems like new cultural geography are not good at analyzing macro cultural phenomena. Some scholars even believe that the study of macro cultural phenomena from the perspective of geography has returned to the traditional cultural geography research framework [2,6,7]. A primary reason behind is that with respect to macro cultural phenomenon, it is difficult for the technologies in the past to obtain data used to express meaning, value, and discourse in a specific social context on a large scale. Yet, those are the very concerns of new cultural geography [8,9]. In recent years, with the rise of large data technology, a large number of new acquisition and analysis approaches have emerged when it comes to social data production, social and cultural data acquisition and analytical methods, providing infinite possibilities and opportunities for the macro level study of new cultural geography [7,10]. Among these social and cultural data resources, the typical ones include the social media data of Twitter, and Weibo and Zhihu. User's opinion on a certain social phenomenon can be extracted from the large amount of social media data, or even the attitudes of specific groups toward daily life can be analyzed [11,12]. The emergence of artificial intelligence and increasingly mature machine learning technologies such as NLP provide more scientific approaches to word analysis, opinion extraction and emotion analysis, based on big data of social text [13-15]. 这 This is no doubt good news for the macro cultural phenomenon research launched under the framework of new cultural geography.

New cultural geography stresses that culture is the product of the proactive construction of different social groups, rather than the crystallization of human civilization as identified by some experts or elites [6,7,10]. An interesting example is that from the perspective of cultural geography, the local and cross-regional differences between South and North China is primarily reflected in the cultural differences in between. From the perspective of new cultural geography, such cultural differences originate from the differences in daily life and practices, which have certain specific meanings and values. The study of how daily life has constructed individual's experience of space and place covers several aspects, such as food, language and custom, etc. [16,17]. The particularities of the diet and customs in South and North China are precisely the carrier of the genuine local culture. In the process of crossing different regions, such as people from North come to South or the other way around, food cultures, dialect cultures are quite prominent. The question is, under the framework of new cultural geography, conflict will take place between local culture and external culture, giving birth to a series of local experience, and label some places with some certain cultural elements [18]. Then how to sense these cultural elements in daily life from bottom to the top?

In the past, cultural geographers are good at analyzing people's impression on places through a large amount of questionnaires or interviews [19,20]. To some extent, the perception of characteristic cultural elements is mainly achieved through interviews or questionnaire data. In the same time, the theory of cultural image is introduced to cultural geography research [21]. Initially as a psychological concept, image was first proposed by American scholar Boulding in 1956 [22]. In 1965, it was introduced by Reynolds to tourism research [23]. Almost at the same time, Kevin Lynch proposed a research method for city image [24,25]. City image is people's topic feeling on the city [26]. In its gradual development process, city image has also been extended from the physical environment of city to cultural image containing non-material elements such as city culture [27]. The two above images belong to the same concept, which is closely related to human thinking. They are the ideas and scenarios of the perceived objects in our mind. From the perspective of culture, cultural image can be treated as cultural symbol, which symbolizes the way the people perceives the culture of something. Using survey and interview data as core material, some scholars explore the connotation of tourism culture from the perspective of cultural image [28,29]. Based on cultural image, some scholars also try to study the genuine identity of ethnic culture [30]. In a word, cultural image theory has been successfully introduced to the study of cultural geography, especially the study of urban geography and tourism geography which emphasizes cultural characteristics.

In fact, the emergence of big data brings new opportunities for research related to cultural image. In recent years, based on various social media data, scholars at home and abroad have conducted many research on cultural image of cities and tourist destinations [31,32]. These social media data is featured by large size of user sample, rapid update frequency. In addition, they contain people's true

emotions. Most of the content comes from the real feelings and opinions of people's daily life and practice. More importantly, all these data contain location information, meaning not only the topic words can be extracted from the massive social media data, for emotion analysis, but also the extracted topic words can be positioned and spatially clustered, in a bid to realize the quantitative, automated and spatialized construction of the map of meaning and comprehensively sense the feedback on various aspects of daily life by people from different regions. What's more, NLP has made great progress in text-based perspective extraction, topic word extraction, and semantic analysis of emotions [33–35]. There are lots of available samples used for place name extraction and type recognition, which provide data and basis methods for analysis in this paper, greatly promotes the feasibility of analysis.

The perception and analysis of cultural image based on data such as social media data complies with the research mode of new cultural geography, which emphasizes the bottom-up cultural meaning and value system stems from people's daily life and practice. This paper adopts the geography-related topic Q&A data in Zhihu, the largest domestic Q&A platform, as main data sources, and tries to construct cognition of cultural differences between South and North China through text analysis, POS tagging, proper nouns extraction, and topic words clustering with NLP as core, and carry out emotion analysis under the influence of cultural differences between South and North China [36–38].

2. Study area and data description

This paper takes the entire territory of China as the study area, as shown in Figure 1. Since the core topic of the study is the interaction and difference of food cultures between South and North China, it is necessary to define north and south from a certain geographical perspective, so as to avoid the ambiguity of the concept of the North and South regions which may affect the analysis results. In China, there is a widely recognized North-south division line which runs through the entire East and West China [39,40], as indicated by the red line in **Figure 1**. As a matter of fact, the publicly recognized division line of south and north is not only substantiated and recognized objectively by geographers, but also found to be topically widely accepted by people topically [41,42].

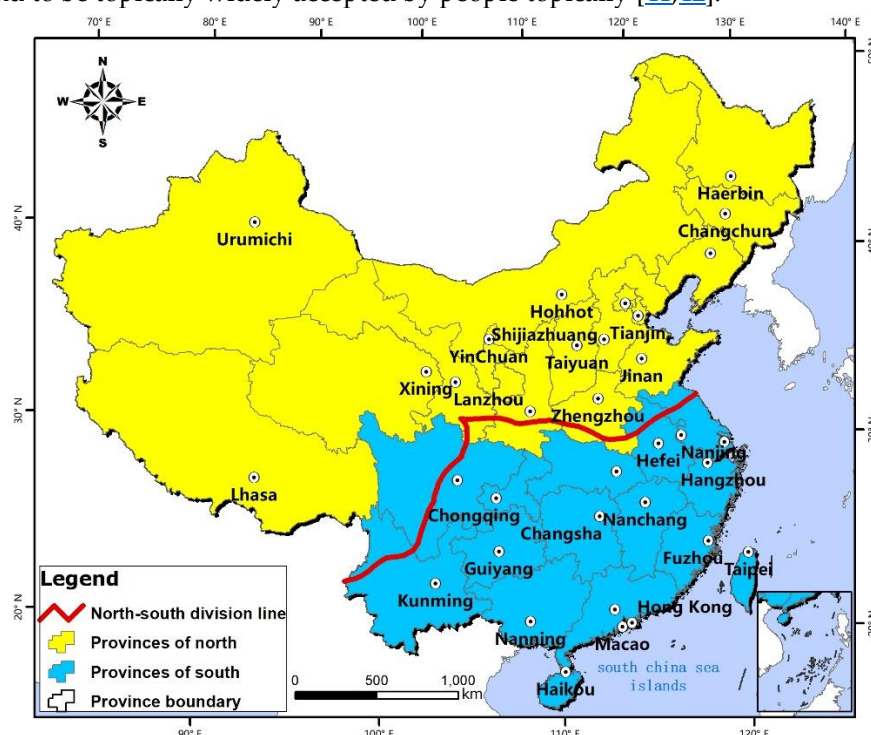


Figure 1. Study area including the North-south division line of China.

The research data of this paper is based on 7,212 user answers to a geographic hot topic on Zhihu, "what are the differences between South and North China that you never know?" Zhihu is currently

the largest knowledge-based network Q&A community in China. According to relevant statistics, as of September 2017, Zhihu has more than 100 million registered users and 26 million active daily users. The time of data collection is February 2018. Since some of the answers were hidden due to violations of the relevant provisions of Zhihu, the actual answer data entries used in this study were 6,128, accounting for 93.23% of the total number of comments.

3. Methodology

This paper mainly includes three parts, including topic data analysis and statistics, spatial modeling and analysis of cultural topic data, and organization of geoscience information map method and its visualized expression. Topic data analysis and statistics mainly cover unsupervised word segmentation, word statistics, corpus training, word classification and statistics, etc. Spatial modeling and analysis of cultural topic data is primarily the analysis of statistics data through GIS spatial analysis and modelling, to explore its spatial distribution characteristics and laws. Finally, using geoscience information map, the cognition hotspots of the culture of south and north, and their hotspot cultural elements are visualized by way of map. And emotion analysis is performed on the words using NLP.

3.1. Text preprocessing

Regionalization In terms of preprocessing, the paper mainly introduces it from two aspects. First, Chinese word segmentation techniques. Second, removal of stop words.

Most of the replies and comments in Zhihu are Chinese-based. Words are the smallest meaningful unit of speech. Different from English words which take space as natural delimiter, Chinese takes characters as basic unit, without distinct distinguishing mark between words. Thus, word segmentation of Chinese is fundamental and essential for emotion analysis. Some of the mature Chinese word segmentation system include ICTCLAS by Chinese Academy of Sciences, IRLAS developed by Harbin Institute of Technology and SCWS developed based on C language, etc. This paper adopts the ICTCLA developed by Chinese Academy of Sciences [43] as such system have the four following features. 1) Using Cascading HMM, ICTCLAS puts Chinese lexical analysis into a unified framework, which improves the accuracy of word segmentation. 2) It supports multi-threaded calls, which improves the speed of word segmentation. 3) It can not only recognize simplified and traditional Chinese, but also supports the recognition of English and symbols. 4) It allows users to add a customized dictionary of emotion to the word segmentation dictionary to make the performance more stable.

Stop words are also know as function words, which is a class of words that have no actual meaning. In this paper, in the semantic analysis of Zhihu's comment text, stop words refer to form words, prepositions, pronouns and other irrelevant characters, with very high or low appearance in the text. Take the frequent words such as "#", "@", "http://" in Zhihu text as an example, these words are all characters irrelevant to the analysis of emotion, and have no practical meaning for the study in this paper. Therefore, it is necessary to filter out the stop words before analyzing the text.

3.2. Spatial interaction patterns between different cultures

(1) Place name relation inference

Place names of China are divided into three levels. Level 1: regions (e.g. Northwest, Southeast region, Jiangsu, Zhejiang, Shanghai Region), level 2: provinces (e.g. Jiangsu, Anhui) and level 3: cities (e.g. Nanjing, Chongqing, Guilin), which are used as our place name pool. Whereas, when talking about South-North differences, the text may involve a large number of place names, and the relationship between place names is complicated. The complex relationship between multiple place names makes the reasoning very challenging. Through the analysis of a large amount of Zhihu text, it is found that there are two main relationships between place names, namely, inclusion relation and parallel relation. Next, taking the real text content of Zhihu as an example, this paper briefly explains how to reason the relationship between place names. E.g. Heilongjiang, Northeast, North in **Table 1**

are in inclusion relation, while Beijing, Xi'an, Suzhou, Nanjing, Xiamen in **Table 2** belong to parallel relation.

Table 1. Sample of extraction of place names with inclusion relation.

Steps	Result
Zhihu text	I'm from Guangdong . What struck me the most when I lived in Heilongjiang was that when I passed a food street while hanging out with a friend from North
Place name extraction	Guangdong, Heilongjiang, North
Place name analysis	The text contains 3 place name information, out of which only one belongs to south, and according to the semantics, the place name is an origin place. The remaining two names related to North are in an inclusion relation (North>Heilongjiang), and based on semantics, belong to destination place. Therefore, according to the principle of accuracy, the place name of the text can be determined as (Guangdong, Heilongjiang). According to the direct preposition "from" between Guangdong and Heilongjiang, it can be inferred that that the Cantonese headed for Heilongjiang.
Final result	(Guangdong → Heilongjiang)

Table 2. Sample of extraction of parallel place names.

Steps	Result
Zhihu text	I'm from Harbin , before I went to Beijing for university, I have never been out of my hometown, which has limited my experience. However, wherever I go, I am used to local food. For example, I had noodles for several days in Xi'an , and I'm still thinking of the noodle that I had at a stall midnight. When I went to Xiamen for the first time, I exclaimed how delicious was the seafood even for me, who barely had any seafood before for 20 years.
Place name extraction	Harbin, Beijing, Xi'an, Xiamen
Place name analysis	The text contains 4 place names, out of which, only one is an origin place. The remaining 3 names are in parallel relations and all belong to destination place. Therefore, it can be inferred that this text is about a Harbin people's comment on the foods of another three cities.
Final result	(Harbin → Beijing), (Harbin → Xi'an), (Harbin → Xiamen)

(2) Classification of hot cultural topics

Topic classification can effectively detect the hidden information behind massive texts, and has important significance for text information detection such as topic detection and text classification. LDA (Latent Dirichlet Allocation) is a document topic model which includes three level, namely, document, topic and word. The model can identify latent topic information in large document sets or corpus using unsupervised machine learning techniques [44,45]. The main idea is to treat each document as a mixed distribution of all topics, and each topic as a probability distribution over the word. The process of generating a text by LDA model is as follows:

- 1) Sampling from the Dirichlet distribution whose hyperparameter is α to generate the topic distribution θ_i of document d_i .
- 2) Perform the following three operations to each word in d_i . First, sampling from the multinomial distribution θ_i that represents topic to generate its corresponding topic Z_{ij} . Second, sampling from the Dirichlet distribution whose hyperparameter is β to generate topic Z_{ij} 's corresponding word distribution $\psi_{Z_{ij}}$. Third, sampling from the multinomial distribution $\psi_{Z_{ij}}$ representing words to generate word W_{ij} .

This paper uses the topic class identified by LDA as the input of emotion analysis of Zhihu user, and employs Topic Coherence index evaluate the topic quality of LDA modeling [46]. Topic

Coherence obtains a score for topic by calculating the semantic similarity between high-score words in a topic, specifically, $coherence(V) = \sum_{(v_i, v_j) \in V} score(v_i, v_j, \epsilon)$. V is the set of words describing a topic, ϵ is a smoothing factor to ensure that the returned score is a real number. In addition, this paper also employs UMass topic coherence evaluation method proposed by Mimno et al. to evaluate topic quality [47]. The equation of UMass score is $score(v_i, v_j, \epsilon) = \log \frac{D(v_i, v_j) + \epsilon}{D(v_j)}$ (2). Where $D(v_i, v_j)$ is the total number of Weibo text containing word v_i and v_j , $D(v_j)$ is the total number of Weibo text containing word v_j . This paper uses Topic Coherence score to determine the most appropriate number of topics [46].

After processing and analysis based on the LDA model, we found that the differences between North and South mainly include staple foods, snacks, salutations, tastes, clothing, accents, weather, festivals, etc., and some of these topics can be merged, such as staple foods, snacks, tastes, etc. can be combined as diet. So on and so forth, the paper divides relevant topics into 15 categories, such as location, geographical object, place name, animal, dialect, costume, climate, emotion, body parts, body characteristics, life, lifestyle, time, diet and plant, as shown in Table 3.

Table 3. Topic classification and statistics.

Category id	Topics	Example	Number of topical words	Ratio
1	location	canteen, dormitory, home	6085	5.65%
2	geography object	mountain, sea, Yangtze river	177	0.16%
3	place name	Nanjing, Shandong, Suzhou	27084	8.10%
4	animal	cockroach, snake, gecko	1450	1.35%
5	dialect	Cantonese, Rhotic accent, accent	2966	2.76%
6	costume	jacket, short sleeve, shirt	1907	1.77%
7	climate	winter, ultracold, muggy	6063	5.63%
8	emotion	like, disgusting, can not stand	6297	5.85%
9	body parts	stomach, skin, mouth	1354	1.26%
10	human figure	fair complexion, tall, fat	534	0.50%
11	living habit	bath, clean, laundry	6474	6.01%
12	living goods	toothbrush, window, heating	3072	2.85%
13	time	spring festival, summer vacation	4373	4.06%
14	food	hot pot, rice, noodles	39548	36.74%
15	plant	flower, grass, tree	257	0.24%

(3) Emotion analysis

After Zhihu's topic is created by the forum administrator, only the netizens interested in the topic will respond or comment, so most of the responses and comments are related to the topic. Correctly extracting the sentiment words in the evaluation is the key to emotion analysis. The dictionary method is one of the most important methods of emotion analysis. The basic idea is to use sentences as research units, and separately perform emotion analysis of each sentence in an article, and then summarize and analyze all sentiment words to determine whether the emotion in the article is positive or not. If positive sentiment outweighs negative sentiment, the comment is determined as positive, otherwise it is negative. In order to make the results more accurate, we score the words in the dictionary. Scores for positive sentiments are set as positive values, such as like, good, willing to try, agree with, etc. Score for negative sentiments are set as negative value, e.g., dislike, tastes awful, hard to understand, horrified, oppose, etc. The value of words showing neutral attitude is set as zero, such as neutral, doesn't care, not interested, etc. In addition, different words show different intensity of sentiment, and the stronger the sentiment, the greater the absolute value.

In the dictionary method, the most important thing is the selection of dictionary of sentiment words and the identification of new words. In order to improve analysis accuracy, this paper adopts BosonNLP_sentiment_score, whose each entry consists of word and score [48]. The words and scores are derived from millions of sentiment annotated data from data sources such as Weibo, news, forums, etc. and have recorded a wide range of non-standardized terms such as new online words. After obtaining the sentiment score of the comments regarding food cultural differences between North and South China on Zhihu, we can get a quadruplet consists of origin, destination, food and score. Origin represents the person's hometown, destination indicates the place which the person commented on, food is the object of evaluation, score is the score of the corresponding sentiment.

4. Results

The analysis results mainly include three parts, including spatial distribution pattern of cultural interaction between North and South, hot cultural topics, and analysis of the emotional characteristics of food cultural differences between North and South. The first part mainly analyzes which areas are more concerned and mentioned by people. The second part hot cultural topics extract topic words using NLP, and find out topics with high frequency topic words, and select the elements related to culture as the hot cultural characteristics. Analysis of emotion characteristics is to classify and analyze the sentiment-related topic words, and analyze the emotion polarity category of each word, so as to quantitatively analyze the influence (positive and negative) degree of the North-South cultural differences on people.

4.1 Spatial distribution pattern of cultural interaction between North and South

Analysis reveals that, the cultural concept of South-North division line is not fixed. It can be both the widely quoted objective South-North division line, which was proposed by China's geographers and widely accepted by people and the subjective division line. For example, Guangdong and Jiangsu all belong to South. Yet some Cantonese believe they are the real Southern, and in their eyes, Jiangsu belongs to North. Another example is, in the mind of the people from Northwest, Shandong located in North China belongs to South. Thus, in the analysis of cultural interaction between South and North and its hotspot flow pattern, this paper performs analysis from two aspects, namely, objective and subjective analysis of the cultural interactions between South and North and its hotspot flow pattern. Objectively, the analysis primarily focuses on the detection of the cultural interaction hotspot flow pattern of the areas on both side of the division line, as shown in Figure 3(a). Subjectively, the analysis focuses on the internal hotspot flow patterns on both sides of the division line, as shown in Figure 2(b) and Figure 2(c).

Figure 2 (a) shows 5 hot cultural interaction zones in the North, which are respectively Beijing at city-level, Henan at province-level, and North, Northwest, North China and Northeast at region-level. Also the figure shows 4 hot zones in the South, which are Jiangsu, Sichuan and Guangdong at province-level, and South at region-level. An interesting finding of such grading and hotspot detection is that most of the hotspot zones in the North are at region-level type, while those in the South are mostly at province-level. The further study of the pattern between hotspot zones in the North and South reveals some interesting laws. Two primary interaction patterns are found, namely, discrete interaction flow pattern and aggregate interaction flow pattern. Some people answered the questions by directly using the general terms of South and North to describe the differences between South and North, but the flow patterns they form are different. E.g. as shown in Figure 2 (a), the orange hotspot representing North forms a relatively discrete hotspot interaction flow pattern with South, while the purple hotspot representing South forms a more aggregated flow pattern with North. For other South and North hotspots, such as the Southeast in the North forms aggregate flow pattern with Guangdong in the South, while Guangdong forms relatively discrete pattern with other spots in the North.

Compared with Figure 2(a), which indicates the objective South-North differences and their cultural interaction hotspot flow pattern, Figure 2(b) and Figure 2(c) shows objective differences and cultural interaction hotspot flow pattern. Figure 2 (b) shows the interaction flow patterns among the

regions north of the division line. Figure 2(c) shows the interaction flow patterns among the regions south of the division line. Figure 2(b) shows that Northeast, which belong to North, forms strong interaction flow pattern with North China interaction flow pattern. And Northwest forms strong interaction pattern with North China and Henan respectively. Compared with Figure 2(b), more hotspot zones are formed in Figure 2(c). What's more interesting is that among all strong hotspot interaction flow patterns, Guangdong has formed a strong hotspot flow pattern with almost all other hotspot zones, which is consistent with the view of Cantonese that except for Guangdong, all regions belong to North.

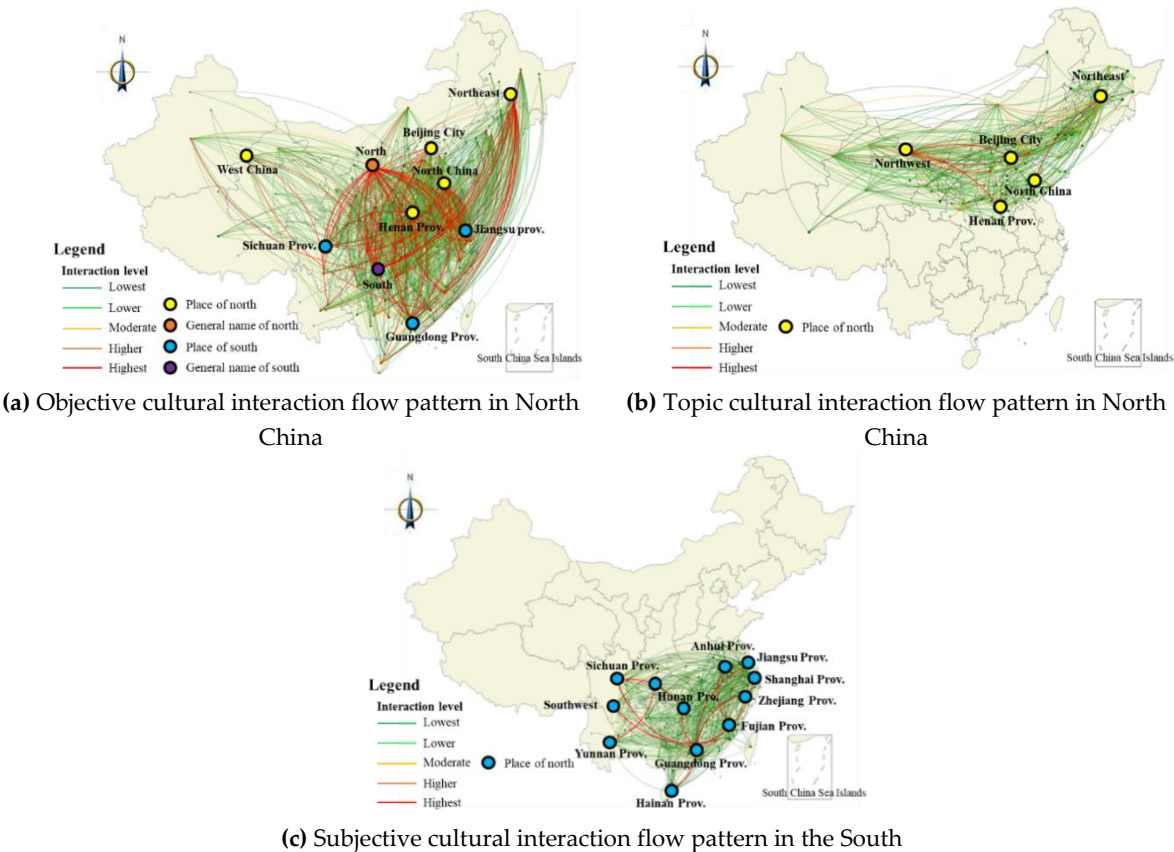


Figure 2. Objective and subjective South and North and its cultural interaction flow pattern.

4.2 Hotspot cultural topics

16 categories of topics were obtained using NLP for topic words extraction. The type number, topic category, word example and the ratio of the number of topic words to the total number of topic words of each topic are shown in Table 3. Some of these topic categories are strongly related to culture, such as diet, dialect, etc. Some are less relevant to culture, such as climate, plants, etc. Next, the paper briefly discusses some topics with higher frequency. Table 3 suggests that diet-related topics strongly associated with culture account for 37% of total topic words. Apparently, the differences in food culture between South and North, which people have profound experience in daily life, are most concerned by people. The second topic with most topic words is palace name, which partly reflects the fact that conversation regarding diet is related to geography. The high frequency of emotional words indicates that people tend to express their attitudes when talking about differences between South and North. In addition, there are many topics regarding lifestyle and climate and environment, meaning the difference in this respect between South and North are also widely concerned by people. The frequency of animal-related topics is higher than that related to plant, which may be because people interact more with animals, leading to people's deeper impression of animals.

4.3 Analysis of emotional characteristics based on cultural differences of food between North and South

Whether the attitude of people from South to the food culture of North, or the other way around, or even the attitude of people from North and South toward their own food culture will show positive, negative or neutral emotions. According to the analysis in the previous section, cultural differences of food between North and South is the most popular topic. Thus, we mainly analyzed the emotional characteristics under the influence of food cultural differences between South and North. The results are shown in Figure 3. We mainly analyzed southerners' attitude towards the food culture in North, and northerners' attitude towards the food culture in South and the attitude of people from both South and North towards their own food culture (marked as others in the Figure). Their respective percentage is shown in Figure 3(b), which shows that people from North and South discussed little about their own food culture, accounting for only 8% of all results. The extraction results of the other two types are mainly analyzed here. Figure 3 (a) and (c) are respectively the ratio diagram of southerner's emotions to food culture of North, and northerner's emotions to food culture of South. Figure3 (a) suggests most southerners show negative emotions towards the food culture of North, accounting for around 67%. Positive and neutral emotions account for similarity of small percentage, which are 17% and 16%, respectively. The percentage of three emotions of northerners towards the food culture in south, as shown in Figure3(c), are similar to Figure 3(a), indicating that when people face a different food culture than their own, they tend to show resistance.

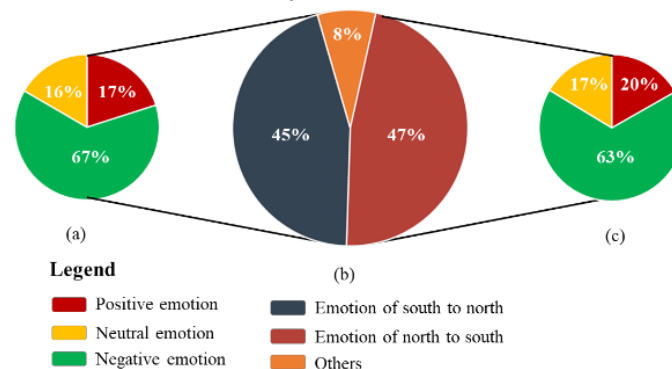


Figure 3. Emotion ratio pie chart under the influence of cultural differences between North and South.

The above analysis only shows the emotional characteristics under the influence of food culture differences, but cannot detect the spatial distribution patterns and relationships of various emotions. Thus, we try to construct emotion flow map, in order to reveal spatial distributions and spatial interaction patterns of different types of emotions through spatial visualization. Figure 4 and Figure 5 show the spatial distributions and interaction patterns of the three emotions of southerners towards the food culture of North, and of northerners towards the food culture of South, respectively. Each emotion flow map consists of several emotion flows, and each emotion flow consists of three elements, including the origin region node of the flow, the destination region node of the flow, and the connection line between the flow. The color of the line indicates the strength of such emotional flow. When referring to South and North, some people are used to adopting the name of the city, while others use the name of the province, and more people directly refer to the regional name for South and North. In the figure, Yellow indicates the province or city in the North while blue indicates the province or city in the South. Orange is the general terms of the entire zone, which is located at the center of the North. Purple is the general term of South. It is located at the center of the entire South. Just as mentioned before, as many people use "North" and "South" to refer to the North and South of China, in each emotion flow map, the flow pattern between North and South are strong.

As can be observed in Figure 4 (a), Shanxi people show strong positive emotion towards the food of Guangdong, other strong positive emotions include Shaanxi's emotion to Jiangsu, and Xi'an's emotion to Jiangsu, etc. Figure 4 (b) shows that, compared with the positive emotions in Figure 4 (a), its coverage is even broader. What's more interesting is that compared with the remaining North, Northeast shows more negative emotions towards the food of South, and many North regions show negative emotions toward foods in Guangdong and Shanghai. Compared with Figure 4(a) and Figure 4 (b), Few northerners have a neutral attitude towards the food culture in the South, as shown in

Figure 4 (c). All in all, whether it is positive emotion or negative emotion, Guangdong is the most concerned province, which not only has strong flow patterns, but also has a wider influence on North, in terms of food culture.

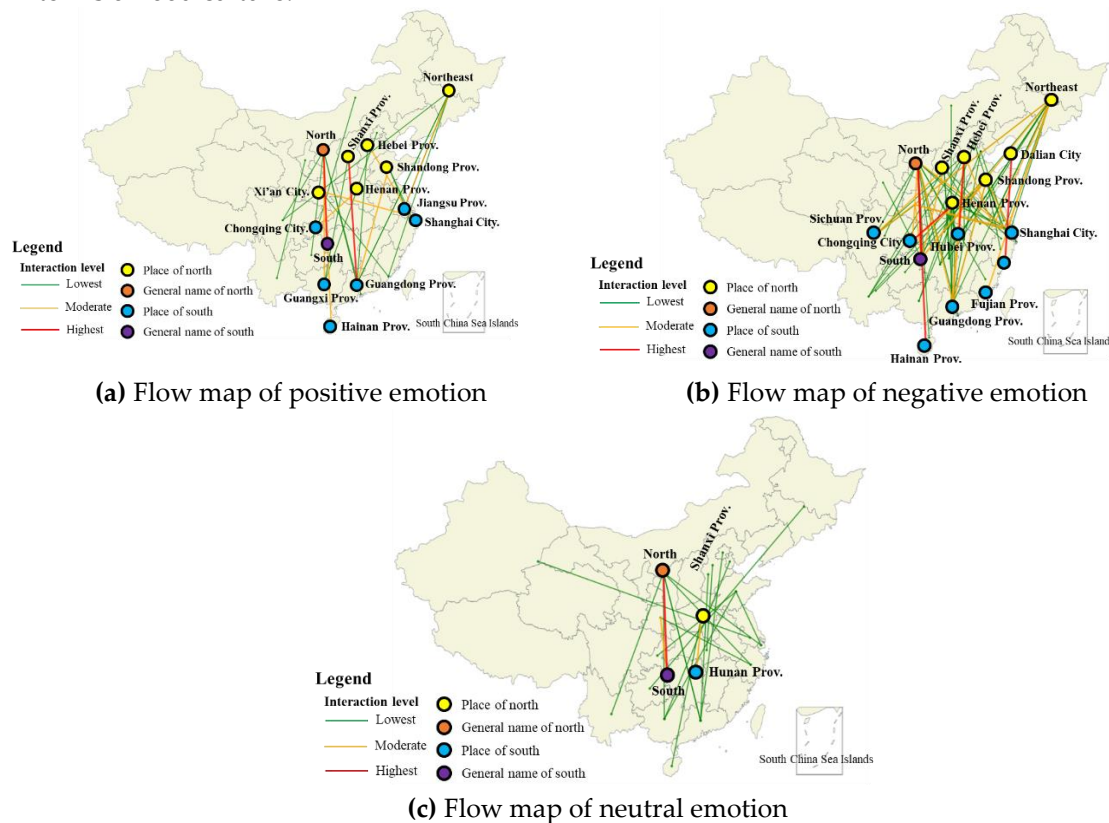


Figure 4. Emotion flow map of people from north to south of China.

As shown in Figure 5(a), apart from the strong positive emotion flow between the generally referred South and North, other pronounced positive emotions are shown by the hinterland area of the South towards the food culture of North, namely central south area, such as Hunan and Hubei, etc. In comparison, there is only a few discussions about southerners' positive attitude towards the food culture in Northwest. Discussion of Northeast also mainly focuses on the Yangtze River Delta, which is near Northeast. Compared with Figure 5 (a), South regions which show negative emotions to the food culture of North are mainly coastal regions such as Shanghai, Zhejiang, Guangdong and Hainan, etc. By comparing Figure 5(a) with Figure 5(b), it can be found that whether the discussions are with positive or negative attitude, such discussions are mainly in Northeast, Inner Mongolia Autonomous Region and Beijing. The main reason for such pattern in Beijing and Northeast is the large number of colleges and universities and developed economy [49,50]. In comparison, Inner Mongolia trails in terms of colleges and universities and economy. However, as Inner Mongolia is a tourist destination with distinctive ethnic characteristics and amazing grassland landscape, such pattern is formed. Figure 5(c) indicates that, except for the emotion flow pattern between the generally referred South and North, Guangdong's neutral attitude flow towards Northeast is strong.



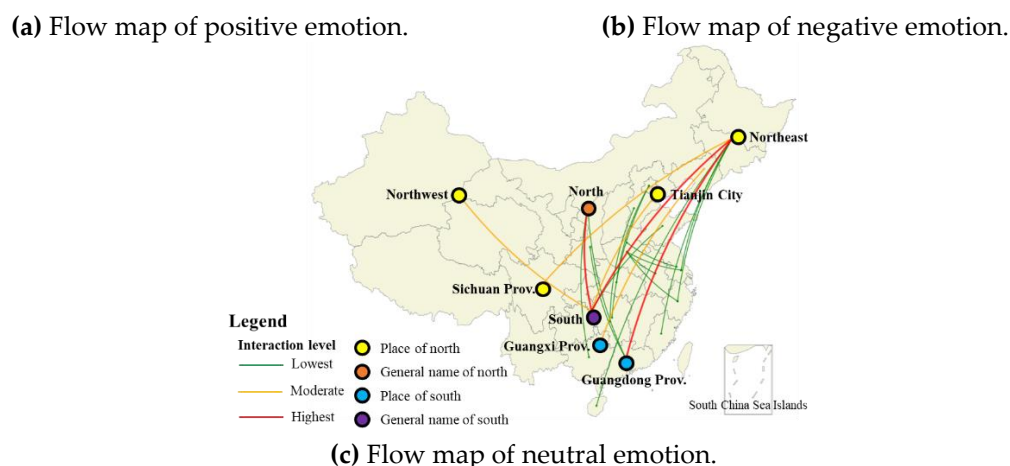


Figure 5. Emotion flow map of people from south to north of China.

5. Discussion and conclusion

5.1. Discussion

(1) Fuzziness of symbol of place names

Although this paper extracts the place names in the text accurately, the diversity and inaccuracy of the expressions by users may lead to some deviations in terms of the scale of the extracted place names. In addition, it is difficult for the current NLP to extract place names in various forms completely and correctly. The two reasons result in the certain fuzziness of symbol of place names extracted in this study. For instance, in describing the differences between South and North, a user from Shenyang (city), Liaoning (province), Northeast China, may use various expressions to refer to his/her residence, such as Shenyang (city), Liaoning (province), Northeast China or even directly North, which may lead to the issue of polysemy and multi-scale for the same place name. In the study in this paper, specific place names are employed to be consistent with the most specific place name symbol used by the respondent.

(2) Precision of semantic parsing

Sometimes, some texts use a place from South to compare with multiple Northern places when expressing the cultural differences between North and South, or vice versa, or even use multiple Northern places to compared with multiple Southern places. The first two situations can be well addressed, using the solution as described in section 3.1, which has provided specific explanations and examples. Regarding the third situation, we have adopted a generalized approach, namely, if multiple places belong to the same administrative area at upper level, then this administrative area at upper level is used instead. Although such treatment has, to some extent, addressed the issue, its mainly drawback is the reduction of location accuracy of place name.

5.2. Conclusion and future direction

This paper takes the cultural differences between South and North China as perception target and takes the 6,128 answers to the question "what are the differences between South and North China that you never know" on Zhihu- the biggest knowledge Q&A platform of China, as main data source. The paper analyzes and extracts all the words in the texts using NLP, with a focus on place names, words related to culture, and emotional words. In addition, the paper collects the statistics and analyzes the hotspot interaction flow patterns, hotspot cultural elements and main emotions of the differences between North and South using clustering. The main conclusions are as below.

In the book the Clash of Civilizations and the Remaking of World Order[51], Samuel Huntington, Professor of International Relations of Harvard University, believes that the reason behind the conflicts between countries and regions in today's world is not the conflicts between religions and economy, but the conflicts between different cultures. The core view of this theory is when people

from different regions have conflicts due to differences in customs, diet habits, etc., they should learn as much as possible about each other's cultures, so as to understand the cultural differences and promote identity recognition. Besides, the interaction and clash of cultures will facilitate the spread and integration of cultures.

The study results of this paper suggest that, firstly, people from South has a fuzzy identification of the cultural boundary of North, and tend to use larger scale to describe the shared cultural characteristics of North. On the contrary, people from North tend to use more specific and areas with smaller scope to define the cultural characteristics of South, which indicates the difference of cognition of each other's culture between people from South and North. Secondly, the analysis of hotspot topic words related to the differences between South and North reveals that the differences of food culture is the most concerned topic, showing that food culture plays an essential role in culture production in people's daily life. Finally, the analysis of people's emotions to each other's food culture reveals that people from both South and North tend to show negative emotions to each other's food culture. This shows that when different cultures interact, in most situations, conflicts will appear rather than appreciation of each other's culture. However, instead of coming into being naturally, cultures are generated under the influence of social environment, meaning that as long as people understand more about a certain culture, they will inevitably adapt to or even accept the culture in a subtle way.

This paper has made efforts to the application of new data of cultural cognition and quantitative analysis method. Yet there are still some shortcomings and it is urgent to carry out research on cultural significance, value judgment and even emotional attachment and identity of relevant cultural elements, from the perspective of new cultural geography. For example, further analysis can specifically tell which characteristics belong to the cultural symbols of North and which belong to South, and detect each emotional word is generated in which cross-region process. All these require further study.

References

1. Lopez, J.J.R. The symbolic appropriated territory, an updated tradition in the new cultural geography. *Rev Geogr Venez.* **2018**, *59*, 434-447.
2. Cresswell, T. New cultural geography - an unfinished project? *Cult Geogr.* **2010**, *17*, 169-174.
3. Jackson, P. New directions in cultural geography revisited. *Area.* **2016**, *48*, 367-370.
4. Mitchell, D. Cultural geography: A critical introduction. *Antipode.* **2002**, *34*, 334-339.
5. Goheen, P.G. Cultural geography: A critical introduction. *Annals Of the Association Of American Geographers.* **2001**, *91*, 572-573.
6. Brazeau, R. The cultural geography of revolution. *Ir Stud Rev.* **2019**, *27*, 442-446.
7. Peach, C. Social geography: New religions and ethnoburbs - contrasts with cultural geography. *Prog Hum Geog.* **2002**, *26*, 252-260.
8. Cook, I. A new vocabulary for cultural-economic geography? *Dialogues Hum Geogr.* **2019**, *9*, 83-87.
9. Zimmerer, K.S. Cultural ecology: At the interface with political ecology - the new geographies of environmental conservation and globalization. *Prog Hum Geog.* **2006**, *30*, 63-78.
10. Steinbrink, M.; Aufenvenne, P. On othering and mainstreamisation of new cultural geography. Some scientometric observations. *Mitt Osterr Geogr G.* **2017**, *159*, 83-104.
11. Turan, Y.; Cekic, A. Coping with loneliness: Loneliness, religious coping, religiosity, life satisfaction and social media usage. *Cumhuriyet Derg.* **2018**, *22*, 1711-1745.
12. Orben, A.; Dienlin, T.; Przybylski, A.K. Social media's enduring effect on adolescent life satisfaction. *P Natl Acad Sci USA.* **2019**, *116*, 10226-10228.
13. Andrade, C.S.; Santos, M.Y. Sentiment analysis with text mining in contexts of big data. *Int J Technol Hum In.* **2017**, *13*, 47-67.
14. Naji, M.; Daoudi, N.; Ajhoun, R. Spatio-temporal context for improving sentiment analysis accuracy. *Int Conf Multimed.* **2018**, 390-395.
15. Majumder, N.; Hazarika, D.; Gelbukh, A.; Cambria, E.; Poria, S. Multimodal sentiment analysis using hierarchical fusion with context modeling. *Knowl-Based Syst.* **2018**, *161*, 124-133.
16. Chera, M. Food culture: Anthropology, linguistics, and food studies. *Soc Anthropol.* **2018**, *26*, 579-581.

17. Eker, G.O. Food culture as a modern world ritual from a different aspect: Implicit functions of being tested/initiation and revenge. *Milli Folklor*. **2018**, 170-183.
18. Edvardsson, B.; Enquist, B.; Johnston, R. Co-creating customer value through hyperreality in the prepurchase service experience. *Journal of service research*. **2005**, 8, 149-161.
19. Kandra, S.; Ghosh, M. Environmental perception: Image based analysis of people's impression of places. *International Journal of Applied Environmental Sciences*. **2017**, 12, 1223-1239.
20. Mirnig, N.; Strasser, E.; Weiss, A.; Tscheligi, M. In *Studies in public places as a means to positively influence people's attitude towards robots*, International Conference on Social Robotics, 2012; Springer: pp 209-218.
21. Portugali, J. On the relations between cognitive maps and images of cities. *Cogn Process*. **2018**, 19, S6-S6.
22. Boulding, K.E. The image: Knowledge in life and society. *Journal of Political Economy*. **1957**, 54.
23. Reynolds, W.H. The role of the consumer in image building. *California Management Review*. **1965**, 7, 69-76.
24. Filomena, G.; Verstegen, J.A.; Manley, E. A computational approach to 'the image of the city'. *Cities*. **2019**, 89, 14-25.
25. Lopez-Pelaez, J.M. The image of the city. *Proy Progreso Arquitect*. **2018**, 154-154.
26. Guo, C.C. The important relationship between environment design and city image. *2018 7th International Conference on Social Science, Education And Humanities Research (Ssehr 2018)*. **2018**, 31-34.
27. Oshimi, D.; Harada, M. Host residents' role in sporting events: The city image perspective. *Sport Manag Rev*. **2019**, 22, 263-275.
28. Li, J.J.; Lu, Y. Chinese culture in tourist research: A review and comparison of chinese and english studies in 1993-2012. *Tour Rev*. **2016**, 71, 118-134.
29. Wong, I.A.; McKercher, B.; Li, X.P. East meets west: Tourist interest in hybrid culture at postcolonial destinations. *Journal Of Travel Research*. **2016**, 55, 628-642.
30. Beriault, M.; Laurier, C.; Guay, J.P. Allegiance to gang culture: The role of ethnic identity among young offenders in the montreal area. *Criminologie*. **2017**, 50, 287-310.
31. Oh, D.E.S. Social life in social media. *Fudan J Hum Soc Sci*. **2018**, 11, 119-122.
32. Hawi, N.S.; Samaha, M. The relations among social media addiction, self-esteem, and life satisfaction in university students. *Soc Sci Comput Rev*. **2017**, 35, 576-586.
33. Ly, T.; Pamer, C.; Dang, O.; Brajovic, S.; Haider, S.; Botsis, T.; Milward, D.; Winter, A.; Lu, S.; Ball, R. Evaluation of natural language processing (nlp) systems to annotate drug product labeling with meddra terminology. *J Biomed Inform*. **2018**, 83, 73-86.
34. Moreno, D.C.; Vargas-Lombardo, M. Design and construction of a nlp based knowledge extraction methodology in the medical domain applied to clinical information. *Healthc Inform Res*. **2018**, 24, 376-380.
35. Purks, J.L.; Harris, M.; Anderson, K.E.; Shoulson, I. Natural language processing (nlp) of verbatim consequence patient-reported outcomes highlight functional ability as important to patients. *Neurotherapeutics*. **2018**, 15, 1192-1192.
36. Xu, G.X.; Meng, Y.T.; Qiu, X.Y.; Yu, Z.H.; Wu, X. Sentiment analysis of comment texts based on bilstm. *Ieee Access*. **2019**, 7, 51522-51532.
37. Basiri, M.E.; Ghasem-Aghaee, N.; Naghsh-Nilchi, A.R. Exploiting reviewers' comment histories for sentiment analysis. *J Inf Sci*. **2014**, 40, 313-328.
38. Liu, L.Z.; Liu, H.; Wang, H.S.; Song, W.; Lu, J.L. The study of sentiment analysis on chinese comment texts. *2014 5th Ieee International Conference on Software Engineering And Service Science (Icse)*. **2014**, 397-401.
39. Qi, W.; Liu, S.H.; Zhao, M.F.; Liu, Z. China's different spatial patterns of population growth based on the "hu line". *Journal Of Geographical Sciences*. **2016**, 26, 1611-1625.
40. Wang, F.H.; Liu, C.L.; Xu, Y.P. Analyzing population density disparity in china with gis-automated regionalization: The hu line revisited. *Chinese Geogr Sci*. **2019**, 29, 541-552.
41. Chen, M.X.; Gong, Y.H.; Li, Y.; Lu, D.D.; Zhang, H. Population distribution and urbanization on both sides of the hu huanyong line: Answering the premier's question. *Journal Of Geographical Sciences*. **2016**, 26, 1593-1610.
42. Hu, Z.C.; Wang, Y.L.; Liu, Y.S.; Long, H.L.; Peng, J. Spatio-temporal patterns of urban-rural development and transformation in east of the "hu huanyong line", china. *Isprs Int J Geo-Inf*. **2016**, 5.
43. Li, X.D.; Zhang, C. Research on enhancing the effectiveness of the chinese text automatic categorization based on ictclas segmentation method. *Int Conf Softw Eng*. **2012**, 267-270.
44. Tajbakhsh, M.S.; Bagherzadeh, J. Semantic knowledge lda with topic vector for recommending hashtags: Twitter use case. *Intell Data Anal*. **2019**, 23, 609-622.

45. Zhang, Y.L.; Eick, C.F. Tracking events in twitter by combining an lda-based approach and a density-contour clustering approach. *Int J Semant Comput.* **2019**, *13*, 87-110.
46. Stevens, K.; Kegelmeyer, P.; Andrzejewski, D.; Buttler, D. In *Exploring topic coherence over many models and many topics*, Conference on Empirical Methods in Natural Language Processing, 2012.
47. Mimno, D.M.; Wallach, H.M.; Talley, E.M.; Leenders, M.; Mccallum, A. In *Optimizing semantic coherence in topic models*, Conference on Empirical Methods in Natural Language Processing, 2011; pp 262-272.
48. Min, K.R.; Ma, C.G.; Zhao, T.M.; Li, H.Y. Bosonnlp: An ensemble approach for word segmentation and pos tagging. *Natural Language Processing And Chinese Computing, Nlpcc 2015.* **2015**, *9362*, 520-526.
49. Hong, W.; Su, Y.S. The effect of institutional proximity in non-local university-industry collaborations: An analysis based on chinese patent data. *Res Policy.* **2013**, *42*, 454-464.
50. Fisch, C.O.; Block, J.H.; Sandner, P.G. Chinese university patents: Quantity, quality, and the role of subsidy programs. *J Technol Transfer.* **2016**, *41*, 60-84.
51. Huntington, S.P. The clash of civilizations: And the remaking of world order. 1998, 25, 158-162.