Landscape assessment of Mountain Qingxiu using internet landscape discourse and LSTM

Weiwei Wang1,2, Shouyun Shen1,2,3, Qiulin Liao1,2, Cuiyi Chen1 and Yao Sun1
1Central South University of Forestry and Technology, Changsha 41004, Hunan, China;
2Engineering Technology Research Center of Big Data for Landscape in Hunan Province, Changsha 41004, Hunan, China
3Email:shenshouyun@sina.com

Abstract. The internet era has provided a wealth of data sources for landscape assessment by updating the methods of information exchange in public. Landscape assessment using internet discourse and computer language can provide guidance for the rational use of landscape and the scientific management of scenic areas. By taking the internet discourse of travels and reviews about Mountain Qingxiu as research objects, this study extracted the landscape assessment objects and analyze the sentiment responds of landscape using LSTM recurrent neural network. And the sentiment characteristics were analyzed by the sentiment-heat IPA quadrant. According to the results of the study, the public's landscape assessment is both discrete and convergent. The landscape assessment of Mountain Qingxiu is mainly positive; There is no regular correlation between the landscape thermal index and the landscape sentiment assessment; The sentiment scores of landscape elements indicates that the plant landscapes and meteorological landscapes are natural landscape elements with high recognition, while the cultural features, buildings, and landscapes in human landscape elements are attractions with potential for development. The sentiment scores of attractions indicates that the botanical landscapes are natural attractions with high recognition, and the architectural and garden landscapes in humanistic attractions are attractions with potential for development. It can be found that the sentiment-based landscape assessment of Mountain Qingxiu using LSTM and internet discourse can effectively complement traditional methods of landscape assessment and provide a better opportunity for the public to participate in the management of landscapes.

1. Introduction
Landscape assessment derives from an interaction between the physical characteristics of landscape and perceptual/judgmental processes of human[1]. Investigating the public's perception of the landscape is an important issue in the landscape assessment. With high level of reliability, perception-based assessments introduce derived perceptual factors and mediating emotional responses into the relationship[2,3]. Questionnaires and surveys are more commonly used in traditional landscape assessment study. The shortcomings of traditional approaches include the limitations of the sample, the passive participation of the public, and the subjective impact of expert assessment. User-generated Content promotes the creation and organization of emerging internet information resources with the development of a new era of information-based, intelligent, personalized, and privatized internet[4]. As an important medium for landscape perception and cognitive communication, social internet sites have produced a large number of online landscape discourse resources.
Landscape discourse is the way in which people, as the subject of landscape cognition, perceive and recognize landscape objects[5]. As the limitations of communication media such as dialogue and books have been broken, the internet era has opened up a wider and more timely dissemination of discourse, and has also provided a large number of data sources for landscape assessment. Scholars have begun to use internet discourse to study environmental air quality perception and assessment[6], satisfaction assessment of park use[7], people's emotional responses in parks[8], and expand research ideas and methods for landscape image research using big data[9]. It has become a new field of landscape assessment by using internet discourse, but research is still in its infancy, and the theories and methods are also being explored.

On the other hand, rich achievements have been accumulated in the field of computer language research. The mature assessment techniques for emotional attitude analysis are keywords, emotional dictionaries, and machine learning. The machine learning can obtain an emotion classification model by learning an emotion training set labeled manually. It has an advantage in processing large mountains of data[10]. Three mature assessment techniques are used for sentiment analysis, Decision Tree, Naive Bayes, Support Vector Machine (SVM), and Recurrent Neural Network (RNN). The Recurrent Neural Network has the ability of continuous learning and error correction.[11] Long Short-Term Memory (LSTM) was born in 1997. It is one of the models with high efficiency and accuracy in recurrent neural networks[12]. LSTM is widely used in text sentiment analysis owing to its powerful abilities of processing and predicting important events with long intervals in time series[13].

In the background of the internet, one of the difficulties in landscape assessment research is how to break the limitations of traditional evaluation methods and conduct reliable and efficient landscape perception research methods using internet resources. This study was conducted to provide a new sentiment-based method for landscape assessment using internet discourse resources. The internet landscape discourse context is taken as the research object, and the LSTM neural network is constructed to carry out the sentiment analysis of the landscape assessment of Mountain Qingxiu. Based on the sentiment-heat IPA quadrant, this paper analyzes the relationship between attention and sentiment responses and summarizes the sentiment attributes of different cognitive objects.

2. Material and methods

2.1. Case study

![Figure 1. Plan of the core attraction of Mountain Qingxiu.](image)

This work used the core attraction of Mountain Qingxiu Scenic Area as a case study area. The total area of Mountain Qingxiu is 13.54 square kilometers, of which the core attraction area reaches 6.43
square kilometers (Figure 1). As the "green lung" of Nanning with more than 6,100 kinds of plants, Mountain Qinxiu plays a very important role in the construction of the ecological, economic, and social environment.

2.2. Methods

The research conducted in 4 steps: firstly, Python is used to capture and preprocess the data of Mountain Qingxiu's travel notes and comments; then, the landscape assessment objects are extracted by the computer and classified into landscape element and attractions by hand; finally, sentiment evaluation is carried out through the LSTM model, and sentiment feature analysis is performed through the heat-sentiment IPA quadrant (Figure 2).

2.2.1. Data acquisition and preprocessing. Firstly, the internet landscape discourse of Mountain Qingxiu is crawled by the Scrapy crawler frame. Comments and travel discourses from 2009 to 2018 were crawled from 5 websites of China: Ctrip Travel, Qunar, Donkey Mom Travel, Baidu Travel, and Sina Blog. A total of 6,181 valid comments and travelogues were obtained after sifting through duplicate and useless data. Secondly, data cleaning is done using third-party libraries in Python (Openpyxl), built-in functions (set ()), and regular expressions. Finally, the participle and lexical annotation are done with the Python open-source framework Jieba participle and the interface package pyltpa for the LTP language cloud platform, respectively. There are 20,239 words with 97,464 times of word frequency, including nouns, adjectives, verbs, and quantifiers. Vocabularies of 174 landscape elements (landscape materials or scenery with independent appreciation value), and 56 attractions (a basic unit of the landscape with relative independence and integrity, composed of associated landscape features) appear in the nouns. According to the scenic features of the attractions, the landscape elements are divided into three categories in the first level and 11 categories in the second level; the attractions are divided into three categories in the first level of and eight categories in the second level.

2.2.2. Network construction, sample training and test of LSTM. LSTM network construction is implemented in three steps: firstly, the model is generated by loading word vectors, then the ID matrix is created for the training set, and finally, the LSTM calculation unit is created. Model training and testing are carried out as the internet is set up. By cutting sentences from the captured text, a total of 23,578 samples were obtained. And 4,700 (20% of the total sample) were labelled manually as positive
(contains adjectives and degree adverbs expressing positive emotions), neutral (factual descriptions only) and negative (contains adjectives and degree adverbs to express negative emotions). A total of 3,760 pieces of data (80% of the 4,700 pieces of data) are used to build the training set for sentiment evaluation machine learning, and the remaining 20% of the data were used as the test set for model training test. After the first training with 100 times of circular training, the test accuracy reached 77.6%; for the second time, the number of circular training was increased to 200 times, and the test accuracy reached 87.4%; At the end, in addition to the initial 4,700 data, 1,500 data items were added to the training set with 200 times of circular training. And the accuracy of the test reached 92.45%.

2.2.3. LSTM sentiment analysis. According to the content of the discourse, a total of 10,470 sentences samples related to landscape elements and attractions were extracted for the aesthetic assessment. Sentiment assessment of landscape was carried out by the LSTM model trained before. And the sentiment intensity was refined by score, that is, 6-10 points represent positive evaluation, 4-6 points represent neutral assessment (excluding 6 points), and 0-4 points represent negative assessment (excluding 4 points).

The calculation formula of sentiment score is as follow:

\[ S = \text{P} \times 10 \]  
(1)

In the formula: \( S \) represents the sentiment score of a single sentence sample, and \( \text{P} \) represents the positive or negative probability of a single sentence sample.

The formula for calculating the sentiment score of a landscape element or attraction are as follow:

\[ SS = \frac{1}{n} \sum_{i=1}^{n} S_i \]  
(2)

In the formula: \( SS \) represents the sentiment score of landscape elements / attractions, and \( S_i \) represents the sentence sample score of the i-th element / attraction.

2.2.4. IPA quadrant analysis of sentiment thermal. The thermal index (H) is applied in the research in order to solve the problem that the number of research samples cannot be strictly controlled, which is generated by the spontaneous formation of public discourse. The thermal index refers to the weight of different items in the discourse, which is calculated by the probability of a particular item appearing in the total sample. The unit of the thermal index is a 10-point scale. The higher the score is, the higher the sensitivity of the project in discourse is. The calculation formula is as follows:

\[ H = \left( \frac{SP}{TS} \right) \times 10 \]  
(3)

In the formula: \( H \) represents the thermal index, \( TS \) represents the total frequency of the sample project, and \( SP \) represents the frequency of the specific project.

According to Pareto's principle, items with the thermal index ranking in the top 80% - 90% were extracted for IPA. The IPA quadrant map is drawn with the thermal index as the horizontal axis, the emotional score as the vertical axis, and the mean value of the thermal index and the emotional score as the origin. It is used for sample heat-emotion evaluation analysis to understand the relationship as well as the distribution characteristics between landscape sensitivity and emotion evaluation.

3. Results

3.1. Overall sentiment assessment of landscape

| Table 1. Statistical table of sentiment assessment. |
|--------------------------------------------------|
| Category                              | Sentiment score | Variance | Positive | Neutral | Negative |
|----------------------------------------|-----------------|----------|----------|---------|----------|
| Sentiment score of landscape elements | 8.11            | 3.89     | 85.41%   | 8.02%   | 6.57%    |
| Sentiment score of attractions        | 7.98            | 3.83     | 84.29%   | 9.41%   | 6.30%    |
| Overall sentiment score               | 8.07            | 3.88     | 85.07%   | 8.44%   | 6.49%    |
The overall sentiment score of the landscape is 8.07, with the sentiment score of landscape elements score higher than that of attractions. The variance of sentiment score is 3.88, with the dispersion of landscape element assessment score slightly higher than that of attraction assessment. The assessment is mainly positive with the similarity number between neutral assessment samples and negative assessment samples (Table 1, Table 2).

### Table 2. Examples of sentiment assessment.

| Sentiment polarity | Examples                                                                 |
|--------------------|--------------------------------------------------------------------------|
| Positive           | The air is very good, and the scenery is very beautiful!                 |
| Neutral            | “Dragon Tower” is the symbol of Mountain Qinxiu.                        |
| Negative           | There are nothing interesting, just ordinary plants, and no scenery is unique. |

### 3.1.1. Analysis of sentiment and thermal index of landscape elements.

According to the assessment of landscape elements, the recognition of plant and weather is generally high among the natural elements of Mountain Qinxiu; among the humanistic elements, the human objects, architecture, garden landscape were landscape elements with better development potential. In Category level I, the thermal index of natural elements is much higher than that of humanistic elements and complex elements, while the sentiment scores of natural elements and human elements are both higher than that of complex elements. In Category level II, the types of highly sensitive landscape elements with a thermal index of greater than 1 include the following types: plant landscape, meteorological landscape, and ground landscape; the landscape elements with a sentiment score greater than 8 include the following types: plant landscape, animal landscape, meteorological landscape, architecture, cultural scenery and garden landscape (Table 3).

### Table 3. Statistical table of sentiment assessment and thermal index on landscape elements.

| Category level | Category level II | Items count | Thermal index | Sentiment score | Variance |
|----------------|-------------------|-------------|---------------|-----------------|----------|
| Natural        | Plant landscape   | 84          | 4.425         | 8.17            | 3.64     |
|                | Animal scene      | 6           | 0.174         | 8.02            | 4.33     |
|                | Meteorological landscape | 10       | 1.928         | 8.63            | 3.04     |
|                | Garden landscape  | 17          | 1.173         | 7.72            | 3.48     |
|                | Waterscape        | 15          | 0.525         | 7.59            | 4.05     |
|                | Subtotal          | 132         | 8.225         | 8.17            | 3.61     |
| Humanistic     | Building          | 15          | 0.479         | 8.41            | 3.24     |
|                | Places of interest| 10          | 0.222         | 7.69            | 3.97     |
|                | Humanities        | 5           | 0.085         | 8.66            | 2.03     |
|                | Ground landscape  | 8           | 0.348         | 8.09            | 4.41     |
|                | Subtotal          | 38          | 1.133         | 8.19            | 3.72     |
| Complex        | General Service   | 1           | 0.283         | 7.08            | 6.17     |
|                | Road and square   | 3           | 0.359         | 7.18            | 7.29     |
|                | Subtotal          | 4           | 0.642         | 7.13            | 6.78     |
| Total          |                   | 174         | 10            | 8.11            | 3.89     |

* Items count refer to the numbers of cognitive objects.

A total of 48 core landscape elements with thermal index ranking in the top 80% are extracted for IPA. The quadrant chart proofs that the aesthetic assessment of landscape elements do not change regularity with the thermal index. According to the distribution of the elements in the four quadrants, the sentiment characteristics of the public's assessment of the landscape elements in Mountain Qinxiu are as follows: there are five landscape elements in the first quadrant, the dominant resources of the
attraction are "air", "plant", "orchid" and "architecture", which lie outside the quadrant with the high sensitivity and recognition; there are seven landscape elements in the fourth quadrant, the landscape elements that need to be improved are "peach blossom," "gate" and "Mountain road", which lie outside the quadrant with high sensitivity but middle level assessment; there are 18 landscape elements in the second quadrant, the landscape elements with better development potential are "tropical plants," "roses," and "sunset", which lie outside the quadrant with the middle level sensitivity but high assessment; there are 18 landscape elements in the third quadrant, the landscape elements that need to be improved are "fresh wind" and "spring", which located in the periphery with the low sensitivity and assessment (Figure 3).

Figure 3. Sentiment-heat quadrant of core landscape elements.

3.1.2. Analysis of sentiment and thermal index of attractions. With a perception rate of 73.68%, 56 attractions are mentioned in the online attractions in all the 76 attractions of Mountain Qinxiu. According to the sentiment assessment results of attractions, the plant landscapes in natural attractions is highly recognized in Mountain Qinxiu; and the buildings, gardens in the humanistic attractions have better development potential. In Category level I, the sensitivity and sentiment scores of natural attractions and humanistic landscape are higher than those of complex attractions. In Category level II, the types of highly sensitive attractions with a thermal index greater than 1 include botanical scenery, waterscape, places of interest and comprehensive services; the types of attractions with a score greater than 8 include botanical scenery, architecture, and garden scenery (Table 4).

A total of 48 core landscape elements with thermal index ranking in the top 80% are extracted for IPA. By combining Bartley's law and considering the clarity of the quadrant graph, 19 core attractions with thermal index ranking at 90% of the money are extracted for heat aesthetic IPA (Figure 4). The quadrant chart proofs that the aesthetic assessment of attractions does not change regularity with the thermal index. According to the distribution of the attractions in the four quadrants, it is shown that the sentiment characteristics of the public's assessment of the attractions in Mountain Qinxiu are as follows: there are four attractions in the first quadrant, the dominant resources of the attractions are "Orchid Garden" and "Cycad Garden", which lie outside the quadrant with the high sensitivity and recognition; there are three attractions in the fourth quadrant, landscape elements that need to be improved are "Dragon tower", "Sky pool", which lie outside the quadrant with high sensitivity but middle level assessment; there are nine attractions in the second quadrant, landscape elements with better development potential are "Friendship gallery", "Mountain lawn" and "Lotus pond", which lie outside the quadrant with the middle level sensitivity but high assessment; there are three attractions in the third quadrant, attractions that need to be improved are "ancient road" and "Dong spring", which located in the periphery with the low sensitivity and assessment.
Table 4. Statistical table of sentiment assessment of and thermal index on attractions.

| Category level         | Items count | Thermal index | Sentiment score | Variance |
|------------------------|-------------|---------------|-----------------|----------|
| Natural attractions    |             |               |                 |          |
| Plant landscape        | 10          | 3.131         | 8.15            | 3.61     |
| Landscape              | 1           | 0.006         | 4.84            | 0.37     |
| Waterscape             | 7           | 1.407         | 7.79            | 3.40     |
| Subtotal               | 18          | 4.544         | 8.03            | 3.58     |
| Humanistic attractions |             |               |                 |          |
| Building               | 14          | 0.794         | 8.38            | 2.73     |
| Places of interest     | 13          | 2.925         | 7.83            | 4.13     |
| Garden landscape       | 5           | 0.566         | 8.41            | 2.32     |
| Subtotal               | 32          | 4.285         | 8.01            | 3.69     |
| Compound attractions   |             |               |                 |          |
| General Service        | 4           | 1.154         | 7.71            | 4.96     |
| Road and square        | 2           | 0.016         | 6.93            | 4.88     |
| Subtotal               | 6           | 1.170         | 7.70            | 4.95     |
| Total                  | 56          | 10            | 7.98            | 3.80     |

*Items count refer to the numbers of cognitive objects.

Figure 4. Sentiment-heat quadrant of core attractions.

4. Discussion

(1) The research shows that the public's landscape evaluation of Mountain Qingxiu is both discrete and convergent, which confirms the viewpoint of environmental cognition commonality and personality coexistence proposed by Kevin Lynch [14]. The total sentiment score reaches 8.07, which indicates that the public gives a relatively high landscape evaluation of Mountain Qinxiu. As the natural landscapes have been identified as advantageous resources, the human landscapes are resources with potential development in Mountain Qinxiu. Although landscape sensitivity reflects the focus of public attention, it is not directly related to the evaluation of landscape emotion. The sensitivity of the natural landscape elements and the attractions indicate the influence of the characteristic plant scenery in Mountain Qinxiu.

(2) With a total of 6,181 online reviews and travel notes about Mountain Qinxiu from the year 2009 to 2018, a sufficient number of samples with large time spans were obtained in the study. However, it is challenging to cover different ages due to the restrictions on the users of the internet platform. In addition, the difficulty of collecting background information has also increased by the information security strategy used in some websites. The limited performance of specific samples in the traditional evaluation process is reduced to a certain extent through the study of Internet landscape discourse.
Scenery cognition is mainly visual but not limited to vision. The general recognition of the "air" proves that people's cognition of scenery is a comprehensive experience of all senses[15], in which smell plays an important role[16,17]. The scores of the attractions "Mountain lawn" and the "Friendship promenade" are higher than those of in traditional research[18], which indicates that the photo samples have limitations in space experience[19].

(3) The study quantified the results of sentiment assessment on a 10-point scale, which made up for the shortcomings of single polarity analysis in the expression of sentiment intensity, but failed to carry out a richer analysis of sentiment types. Reference 20 conducts a comprehensive evaluation by combining the sentiment level (positive, negative, neutral) and sentiment categories (anger, disgust, fear, joy, sadness, surprise, and neutrality), which helps further the analysis of the assessor's emotional characteristics[20]. In terms of evaluation technology, the LSTM internet has a high accuracy rate but needs to be further improved in efficiency. The accuracy rate of sentiment evaluation by LSTM internet is 92.4% in this study, and it scores 93.27% in related studies[21], while the accuracy rate of Bayesian classification model is 90.05%[22]. The accuracy of both machine learning models can be improved by increasing the times of training, but with the disadvantage of a large amount of corpus required for long-term learning. Text analysis software or platforms such as ROST CM and API used in some studies have the functions of integrating word frequency analysis, internet analysis, and sentiment evaluation[20,23], which is more convenient to use.

5. Conclusion

(1) This study provided a sentiment-based landscape assessment approach using internet resources. The approach reduced the limitations of the sample in traditional evaluations due to the relatively large sample size obtained for landscape sentiment analysis using internet landscape discourses. The methodology significantly reduces the subjective influence of experts in the assessment. The study conducted a fine-grained analysis on the basis of discourse characteristics, and summarized in depth the emotional characteristics of landscape elements and attractions cognition, complementing the lack of landscape specific analysis in the whole text of existing studies. With relatively high accuracy, reliability and efficiency, the combination of LSTM network with thermal index and IPA quadrant analysis are well adapted to the spontaneous nature of landscape discourse.

(2) Subsequent research should select appropriate discourse samples to analyze the discourse differences of evaluators at different times, regions, generations, genders, and occupations; analyze the study of the sentiment characteristics of discourse in different contexts, and draw on relevant studies to explore the verbal characteristics of discourse in depth; and improve the accuracy and intelligence of computer language processing, increase the comparative study of different intelligent evaluation methods, and determine the best research technology based on the characteristics of the research object.

(3) Although there are still some limitations, on the premise of solving the copyright issues and protecting privacy, the way of sentiment-based landscape assessment using computer language and internet landscape discourse could make up for the shortcomings of traditional assessment methods and provide a broader platform for the public to participate in the management of landscape.

Acknowledgement
Support for this study is provided by Excellent Youth Project of Hunan Provincial Department of Education in China(19B580), and Postgraduate Research and Innovation Project of Hunan Provincial Department of Education in China (CX20190617). In addition, I wish to thank Yan Xiaobin for his great assistance with data processing.

References
[1] Terry C D 2001 Whither scenic beauty? Visual landscape quality assessment in the 21st century Landscape and Urban Planning 54(1-4) 267-281
[2] Ulrich R.1983 Aesthetic and Affective Response to Natural Environment Behavior and the
Natural Environment 6 85-125

[3] Zube E H 1974 Cross-disciplinary and intermode agreement on the description and evaluation of landscape resources Environmental Psychology and Nonverbal Behavior 6(1) 69-89

[4] Li J Y, Zhang Y Y 2017 Big data leads research on tourists' emotional experience Tourism Journal 32(09) 8-9

[5] Wang W W 2019 Theory Building and Empirical Research on Landscape Discourse Assessment Central South University of Forestry and Technology.

[6] Yugu T, Feng Z, Chunyun S, et al. 2019 Social Media Data-Based Sentiment Analysis of Tourists Air Quality Perceptions Sustainability 11(18) 5070

[7] Wang Z F, Zhao J N, Peng Y Y, et al. 2019 Comparative Evaluation of Guangzhou City Parks: Text Analysis Based on Social Media Data Landscape Architecture 26(08) 89-94

[8] Plunz R A, Zhou, Y J , Vintimilla M I C 2019 Twitter sentiment in New York City parks as measure of well-being Landscape and Urban Planning 9(189) 235-246

[9] Cao Y H, Yang P F, Long Y 2019 The Innovation of City Image Cognitive Method Based on Deep Learning--A Case Study of Chongqing Main District Chinese Landscape 35(12) 90-95

[10] Sailunaz K, Dhalwal M, Rokne J, et al. 2018 Emotion detection from text and speech: a survey Social Network Analysis and Mining 8(1) 28

[11] Yan J C, Zhao Z H, Zhao R 2019 Research on Social Media Text Sentiment Analysis Based on Machine Learning China Computer & Communication 31(20) 44-47

[12] Sepp H, Jürgen S 1997 Long Short-Term Memory Neural Computation 9 1735-1780

[13] Yuan T T, Yang W Z, Zhong L J, et al. 2019 Personality-based Microblog sentiment analysis model PLSTM Application Research of Computers 1 1-6

[14] Lynch K 1960 The Image of the City Cambridge MA:MIT Press 35-67

[15] Zube E H 1984 Themes in landscape assessment theory Landscape Journal 3(2) 104-110

[16] Zhao J W, Huang Y D, Wu H, et al 2018 Olfactory effect on landscape preference Environmental Engineering and Management 17(6) 1483-1489

[17] Zhou N D, Li Y, Xu X L, et al. 2018 Based on IPA method research on the influencing factors of tourists' perception of values in rural ecotourism: taking an ecological village of Longyan in Zhongshan County, Guangxi province as an example Journal of Central South University of Forestry & Technology 38(12) 142-146

[18] Zhang M H 2010 The Research on Landscape Evaluation of Qingxiu shan Scenic in Nanning Central South University of Forestry and Technology, Changsha

[19] James F P 2001 Rating reliability and representation validity in scenic landscape assessments[J]. Landscape and Urban Planning 54 149-161

[20] Sailunaz K, Alhajj R 2019 Emotion and sentiment analysis from Twitter text Journal of Computational Science 36

[21] Zhou N, He R Q 2019 Comprehensive Evaluation of Cultural Variety Shows Based on Text Emotion Analysis--Taking CCTV Cultural Variety Show "National Treasure" as an Example Journal of South-Central University for Nationalities( Humanities and Social Sciences) 39(05) 175-180.

[22] Deng Y H 2019 Research on Zhangjiajie Scenic Evaluation Based on Scenic Discourse Central South University of Forestry and Technology, Changsha

[23] Li M Y, Cao Z J, Dang A R, et al 2018 Tourist Flow Analysis and Emotional Detection of Smart Scenic Area Based on Social Network Data--A Case Study of Huangshan Scenic Area Chinese Landscape 34(12) 52-56