


Article

Exploring the Potential of BERT-BiLSTM-CRF and the Attention Mechanism in Building a Tourism Knowledge Graph

Hongsheng Xu ^{1,2,*} , Ganglong Fan ^{1,2}, Guofang Kuang ^{1,2} and Chuqiao Wang ^{1,2}

¹ College of Electronic Commerce, Luoyang Normal University, Luoyang 471934, China

² Henan Key Laboratory for Big Data Processing & Analytics of Electronic Commerce, Luoyang Normal University, Luoyang 471934, China

* Correspondence: xhsls@lynu.edu.cn

Abstract: As an important infrastructure in the era of big data, the knowledge graph can integrate and manage data resources. Therefore, the construction of tourism knowledge graphs with wide coverage and of high quality in terms of information from the perspective of tourists' needs is an effective solution to the problem of information clutter in the tourism field. This paper first analyzes the current state of domestic and international research on constructing tourism knowledge graphs and highlights the problems associated with constructing knowledge graphs, which are that they are time-consuming, laborious and have a single function. In order to make up for these shortcomings, this paper proposes a set of systematic methods to build a tourism knowledge graph. This method integrates the BiLSTM and BERT models and combines these with the attention mechanism. The steps of this methods are as follows: First, data preprocessing is carried out by word segmentation and removing stop words; second, after extracting the features and vectorization of the words, the cosine similarity method is used to classify the tourism text, with the text classification based on naive Bayes being compared through experiments; third, the popular tourism words are obtained through the popularity analysis model. This paper proposes two models to obtain popular words: One is a multi-dimensional tourism product popularity analysis model based on principal component analysis; the other is a popularity analysis model based on emotion analysis; fourth, this paper uses the BiLSTM-CRF model to identify entities and the cosine similarity method to predict the relationship between entities so as to extract high-quality tourism knowledge triplets. In order to improve the effect of entity recognition, this paper proposes entity recognition based on the BiLSTM-LPT and BiLSTM-Hanlp models. The experimental results show that the model can effectively improve the efficiency of entity recognition; finally, a high-quality tourism knowledge was imported into the Neo4j graphic database to build a tourism knowledge graph.

Keywords: tourism knowledge graph; word segmentation; text classification; popular words; knowledge extraction; entity recognition



Citation: Xu, H.; Fan, G.; Kuang, G.; Wang, C. Exploring the Potential of BERT-BiLSTM-CRF and the Attention Mechanism in Building a Tourism Knowledge Graph. *Electronics* **2023**, *12*, 1010. <https://doi.org/10.3390/electronics12041010>

Academic Editors: Mohammed Amin Almaiah and Nathir Obeidat

Received: 15 January 2023

Revised: 14 February 2023

Accepted: 15 February 2023

Published: 17 February 2023



Copyright: © 2023 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

In recent years, the rapid development of Chinese tourism has become the most active industry in the economic industry. According to statistics, from 2011 to 2018, China's total domestic tourism reception and revenue grew at an average annual rate of 13.7% and 20.7%, respectively. In 2019, the combined contribution of tourism to GDP was 10.94 trillion RMB, accounting for 11.05% of total GDP. With the continuous development of internet technology, the tourism market has exploded in terms of data on the internet. The tourism market has generated a massive amount of data on the internet, including many tourism entities such as scenic spots, hotels, restaurants, shopping places and so on.

In order to grasp the future trends of the tourism market and tourists' consumption behaviors and demands, we can use association rule mining, deep learning and knowledge map methods to conduct professional analyses. Tourists can systematically understand

the tourist destination and lay a good foundation for a comfortable tourism experience by viewing the correlation between tourism elements on the tourism knowledge graph.

Online tourism comments are a public resource. Tourists express their opinions freely through online comment platforms, which also leads to a large number of useless and repeated garbage corpora. The knowledge graph can structure heterogeneous knowledge in the field, build associations between difference sources of knowledge, easily realize the joint analysis of complex data and enable users to retrieve information more directly, conveniently and accurately [1]. Therefore, tourism knowledge graphs are an important structured knowledge base for tourism data mining and knowledge storage in the era of big data. In this paper, we first analyze the current state of domestic and international research on tourism knowledge graphs.

JIA Zhonghao et al. proposed integrating the network embedding method into the feature extraction method and established a scenic spots recommendation system based on a tourism knowledge graph [2]. The authors of a different paper took tourism websites as the main data source, carried out the entity extraction and entity alignment of the heterogeneous data and constructed a Chinese tourism knowledge graph [3]. In order to construct the tourism knowledge graph, the authors proposed the GRU_ATT distantly supervised relation extraction model for text extraction [4]. Elsewhere, a translation-based method was used to train the knowledge graph so as to realize the knowledge vectorization of the knowledge graph and finally establish a domain tourism knowledge graph [5]. The authors of another paper constructed a knowledge graph of tourist scenic spots and combined the traditional question-answering model and the fine-grained knowledge graph question-answering model based on BiLSTM+CRF to realize the knowledge question-answering system [6]. The authors used a method based on template matching to construct a knowledge graph of Inner Mongolia tourist attractions, achieving good results [7].

In a further paper, the BERT-BiLSTM model was used to construct a tourism knowledge graph of national tourism information, and the final experimental results proved the effectiveness of the model [8]. Zhang et al. [9] mined tourism knowledge from tourism graphs and used the PCNN model in deep learning to build a tourism knowledge graph. Through comparative experimental analysis, this method can improve the precision and recall rate. Elsewhere, two ontology construction methods were integrated to complete the establishment of a tourism knowledge graph [10]. The authors also used deep learning methods to construct an intelligent question-answering model based on a knowledge graph, which illustrates the effectiveness of deep learning models in the application of knowledge graphs. Other authors demonstrated an improvement in terms of CNN in named entity recognition tasks by adding a CNN structure to the front end of an LSTM [11]. Jianyong DUAN et al. achieved good results on new word detection using BiLSTM+CRF [12]. Elsewhere, a new extension model for medical text personification was proposed. This model starts with the preprocessing of medical text and uses a combination of different text vector representations. The proposed model solves the problem efficiently [13].

The authors of another paper proposed the introduction of a gating mechanism between two layers of GCNs to balance the contextual information and graph embedding information obtained by Bert. The final experiment results in terms of short text classification achieved good results [14]. Li et al. proposed a hybrid model of a neural network architecture with a gated attention mechanism and an attention-guided classifier based on regular expressions [15]. This model can effectively improve the efficiency of text classification by placing high-weight words in the key parts of sentences. In Chinese text classification, the BERT model can dynamically optimize the word vector for specific tasks and extract the contextual relationship between words [16]. Therefore, this model has been widely adopted in various text classification problems.

However, the functions of these tourism knowledge graphs are too narrow, their costs in terms of manpower and time in the construction process is high, and the effects of their knowledge graph representations are insufficient. Therefore, the main aim of this paper was to build a high-quality tourism knowledge graph system. The system can

help tourists accurately and quickly obtain the tourism information they want from the numerous and complicated data on the network and upgrade the tourism industry from a tourism information service to a tourism knowledge service. This paper offers a complete set of methods to construct a tourism knowledge graph. The following is a summary of the rest of this paper.

This paper first analyzes the related technologies, including BiLSTM and BERT; secondly, Chinese word segmentation and the removal of stop words are carried out for tourism texts; third, following the vectorization of tourism words, the cosine similarity method is used to classify tourism texts. The classification effect of the naive Bayes method is compared via experiments, proving that classification based on the cosine similarity method is more effective. In order to accurately obtain popular tourism words, this paper defines the popularity value of tourism products through multiple dimensions and constructs the popularity analysis model of tourism products with emotional analysis. For the popular tourism words obtained, this paper proposes the extraction of knowledge based on BiLSTM-CRF. The knowledge fusion based on the BERT model is carried out to obtain knowledge graph triplet data. In this paper, LPT and Hanlp are integrated into BiLSTM-CRF to build BiLSTM-LPT and BiLSTM-Hanlp models. The experiment shows that the model can significantly improve the efficiency of entity recognition. Finally, standard data is loaded into the Neo4j graphic database, and py2neo is used to build the tourism knowledge graph.

2. Relevant Technology Analysis

2.1. BiLSTM

This paper analyzes the characteristics of BiLSTM and illustrates the role of the BiLSTM model in constructing a tourism knowledge graph. The design features of the long short-term memory network model are suitable for modeling text data [17]. LSTM sets up three gate structures of input, forgetting and output to control the transmission, forgetting the less important information and retaining the information that needs to be remembered for a long period, which, to a certain extent, solves the long dependency problem. LSTM is able to deal with long dependency problems, mainly due to the introduction of self-cycling weights that can be used internally [18]. The output of the implicit layer goes to the output and the implicit layer the next time, so the information can be retained continuously. According to the previous state, the model can deduce the later state and has stronger memory [19]. The formula is as follows:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (1)$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2)$$

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (3)$$

LSTM can learn the historical information above but cannot contact the information below, thus reducing the effect of named entity recognition. Therefore, this paper adopts the BiLSTM model [20,21]. The training sequence is input into the forward propagation LSTM network model, and the forward characteristic information, \overleftarrow{h}_t , is obtained through a forward propagation calculation [22]. The LSTM network model with input back transmission obtains the back characteristic information, \overrightarrow{h}_t , through a back propagation calculation. Then, the forward feature information, \overleftarrow{h}_t , and the backward feature information, \overrightarrow{h}_t , are combined to obtain the final hidden state, H . This combines the forward and backward bidirectional semantic features. The formula is as follows.

$$H = \begin{bmatrix} \overrightarrow{h}_t; \overleftarrow{h}_t \end{bmatrix} \in R^m \quad (4)$$

2.2. BERT

In this paper, the BERT model is used for the purpose of tourism knowledge fusion. The BERT model is a bi-directional encoder constructed by overlaying transformer networks. The transformer network is a feature extraction network based on the self-attention mechanism [23]. Because the transformer network contains multiple encoders and decoders and uses multiple attention layers and residual mechanisms, it supports the parallel processing of sequence data and has a higher efficiency and performance than LSTM [24].

The upper and lower sentences generated from BERT indicate that they pass through the full connection layer, so the self-attention mechanism is integrated after the full connection layer. Self-attention can focus on all the words in the question and help the encoder to better represent the word. The formula for calculating attention is:

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

The core idea of MLM is to randomly select 15% words in the text to mask during model training. After selecting the words to be masked, 80% of the selected words are replaced with [Mask] symbols. Then, 10% of the words are replaced by other words, and 10% of the words remain the same. Since only 15% of these words are predicted for each training, the convergence rate of the model can be significantly reduced. The pre training method of NSP selects the previous sentence, A, and the next sentence, B, for each pre training corpus. In 40% of the cases in the corpus, B is the correct next sentence of A. In all, 60% of the next sentences are randomly generated from the corpus [25].

The multi-head self-attention model enhances the attention capability of the model and is calculated as shown below.

$$MultiHead(Q, K, V) = \text{Concat}(head_1, \dots, head_m)W^O \quad (6)$$

$$head_j = \text{Attention}(QW_j^Q, KW_j^K, VW_j^V) \quad (7)$$

In the above formula, QW_j^Q , KW_j^K , VW_j^V represents the weight matrix of the j th head, m represents the number of heads and W^O is the additional weight matrix. Finally, the matrix is spliced together to obtain the multi head attention calculation result of the j th character in the text. The task of tourism named entity recognition and relation extraction is carried out by using the word feature representation self-supervised on the tourism corpus by the BERT model.

3. Data Preprocessing

The raw data used in this paper were obtained using crawling techniques on encyclopedic and vertical travel websites. More than 5000 pieces of tourism text information were collected, and the collected tourism information includes: attractions, cities, hotels, specialties, food, shopping malls and other aspects. These text types mainly include scenic spot introductions, strategies and evaluations, etc. These texts are not directly applicable to the model learning, the data set had to be preprocessed, including text cleaning, text segmentation and the removal of stop words (including punctuation, numbers, words and other meaningless words). The feature of variable text content leads to a significant amount of unnecessary content, such as noise, text duplication and text redundancy, etc. If the text content is not processed, the classification effect of the classifier will be affected.

3.1. Text Word Segmentation

Chinese word segmentation is the basic function of Chinese human-computer natural language interaction [26]. In natural language processing, word segmentation is usually required first, and the effect of word segmentation will directly affect the effect aspects of speech and syntax tree construction. Text segmentation is an indispensable operation in the

preprocessing process because text classification needs to use words in the text to represent the text. This paper provides a comparative analysis of word segmentation tools:

(1) The word segmentation of Jieba

Jieba is at present the best Python Chinese word segmentation component. It has the following three features: it supports three-word segmentation modes: a precise mode, a full mode and a search engine mode; it supports traditional word segmentation; and it has support for custom dictionaries.

(2) Tsinghua University's THULAC

THULAC features Chinese word segmentation and aspects of speech tagging functions. THULAC has the following characteristics: 1. A strong ability: It is trained by using the largest Chinese corpus (approximately 58 million words) of artificial word segmentation in the world, can tag of parts of speech and has a strong model tagging ability; 2. It has a high accuracy; and 3. it is fast.

(3) The word segmentation of FoolNLTK

FoolNLTK is a Chinese word segmentation tool based on deep learning. Its characteristics are that it has a high accuracy and it is trained based on the BiLSTM model.

(4) The word segmentation of HanLP

HanLP is an NLP toolkit composed of a selection of models and algorithms. Its goal was to popularize the application of natural language processing in the production environment. The main functions of HanLP include: word segmentation, tagging, keyword extraction, automatic summarization and named entity recognition, etc.

(5) The word segmentation of the Chinese Academy of Sciences, NLPPIR

The NLPPIR system supports multiple codes, operating systems, development languages and platforms. The main functions of NLPPIR include Chinese word segmentation, English word segmentation, word tagging, named entity recognition, new word recognition and support for users' professional dictionaries.

(6) The word segmentation of LTP

LTP is an open Chinese natural language processing system. Its main functions include Chinese word segmentation, the tagging of parts of speech, named entity recognition, dependency parsing and semantic role tagging, etc.

In order to test which word segmentation method was more suitable for the word segmentation task in this paper, the above six word segmentation methods were tested. The test results are as follows:

- (1) In terms of time, the same text was tested twice. The times of the six word segmentation methods is shown in Table 1.
- (2) In terms of accuracy and effect of word segmentation, FoolNLTK easily combines interference word segmentation into one word and Thulac makes mistakes in person name segmentation. However, HanLP's word segmentation granularity is relatively large, while FoolNLTK's word segmentation granularity is relatively small, resulting in a large error in word segmentation. Hanlp, meanwhile, is better at handling the names of organizations.
- (3) In terms of installation and code migration capabilities, the comparison results are as follows: Jieba's installation is simple. There are three modes of Jieba, which support a wide range of languages, have a high popularity and feature a better way to operate files. In terms of code migration, it is easy to operate and has strong universality. Although the installation of NLPPIR's word segmentation and LTP's word segmentation are not difficult, they are a little cumbersome compared with the installation of Jieba, and their code migration is relatively weak.
- (4) According to the characteristics of tourism text, Jieba's precise mode can segment sentences most precisely, which is suitable for tourism text analysis; Jieba's search

engine mode can segment long words again, which can effectively improve the recall rate compared with other word segmentation methods. According to the word segmentation rate, Jieba is faster than other word segmentation methods; Jieba includes a variety of functions such as word segmentation, word annotation and named entity recognition, which are embedded in Python. Jieba is the most commonly used tool for word segmentation in Python. Therefore, compared with other word segmentation tools, Jieba was more suitable for constructing the tourism knowledge graph in this paper.

Table 1. Test time of different word segmentation methods.

Word Segmentation Tool	First Time	Second Time
Jieba	0.019	1.878
THULAC	12.119	8.698
FoolNLTK	2.259	2.592
HanLP	3.799	1.843
NLPIR	0.008	0.032
LTP	0.086	0.072

Based on the above comparison results, we selected the exact Jieba segmentation model for tourism article classification. The specific operation steps are as follows: (1) Introduce the open-source word segmentation tool, the Jieba Word Segmentation Toolkit, for word segmentation; (2) import all data; and (3) merge the title and article of the tourism text and subject each merged corpus to Jieba segmentation. In order to prevent the required keywords from being segmented, theme keywords similar to “tourism”, “activities”, “festivals” and other words are generated into user-defined dictionaries.

To illustrate the Jieba word segmentation process, consider the following sentence: “I go to park for tourism”. Figure 1 is as an example of the segmentation of a word by Jieba. When calculating the maximum probability path, the Jieba word segmentation is calculated from back to front.

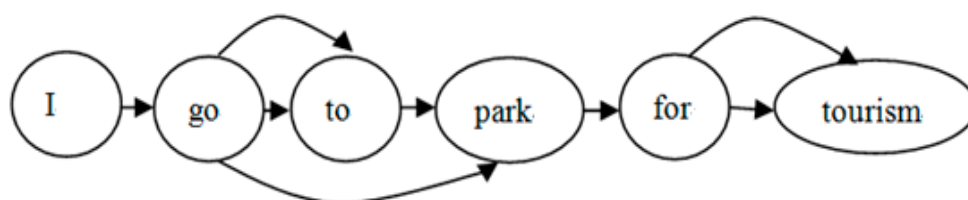


Figure 1. Example of the segmentation of a word by Jieba.

3.2. Remove Stop Words

Removing stop words is also an important part of the preprocessing process. Because not every word or character in the text can represent the text, such as “this”, “month”, “day”, “one, two, three, four”, “I, you and he” or “0, 12... 9”, these words are large in number and have no practical effect. The Baidu and Harbin Institute of Technology stop word lists are currently the most commonly used stop word lists to remove stop words in Chinese. They cover 1335 characters in total, including English characters, numbers and punctuation marks, etc. The process behind removing stop words in this paper was: We screened the Chinese stop words in the authoritative website and used the Baidu and Harbin Institute of Technology stop words lists to filter stop words. Combined with the word segmentation results, the final word segmentation results were obtained after constantly modifying the deactivated word list. In addition, words with high frequency and no contribution were also added to the disabled vocabulary. Table 2 shows the contents of the disabled vocabulary used in this article.

Table 2. Stop word list.

No.	Stop Word
1	This
2	That
3	There
4	Year
5	Month
6	Day
7	Those
8	The
9	By
10	Of

4. Tourism Text Classification

For the collected travel texts, we used two methods for text classification. One is the cosine similarity method and the other is the naive Bayes model. Finally, this chapter compares the classification effects of the two methods on tourism text from five aspects.

4.1. Classification Methods for Tourism Texts

4.1.1. Tourism Text Classification Based on Cosine Similarity

1. Text feature extraction

Text feature extraction is an important part of text classification. The words in the text represent the document in the form of probability. The higher the probability, the better the word can represent the document; otherwise, this word cannot represent the document.

This paper uses the TF-IDF algorithm to extract text features. TF-IDF is a statistical method used to evaluate the importance of a word in a file set or corpus [27]. Its main idea is: if a word appears frequently in one article and rarely in other articles, it is considered that the term has good ability to distinguish and classify, so it can represent the article. It can be seen that the importance of a word increases proportionally with the number of times it appears in the file but decreases inversely with the frequency of its appearance in the corpus. Since $TF \times IDF = TF \times IDF$, the high-frequency words in a specific file and the low file frequency of the words in the entire file set can produce TF-IDF with a high weight. The calculation formula is as follows:

$$TF \times TDF_i = \frac{\sum_{i=1}^n d_{ijk}}{\sum_{i=1}^n d_{ij}} \times \text{WeightingFactor}(T_k) \quad (8)$$

$$d_{ijk} = tf_{ijk} \times \log_{10}(\frac{N}{df_{jk}} \times w_j) \quad (9)$$

$$d_{ij} = tf_{ij} \times \log_{10}(\frac{N}{df_j} \times w_j) \quad (10)$$

$$\text{WeightingFactor}(T_k) = \frac{\log_{10} \frac{N}{df_k}}{\log_{10} N} \quad (11)$$

The formula is described as follows: N represents the total number of files, and w_j represents the number of files without vocabulary j . d_{ijk} is determined by the number of simultaneous occurrences of the word T_k and T_j and the number of files that do not exist in a file. tf_{ijk} indicates the number of times the words j and k appear together in file i . df_{jk} represents the total number of files where the terms j and k occur simultaneously. $\text{WeightingFactor}(T_k)$ indicates the particularity of the word T_k to files. The more common the word T_k is, the smaller the value of $\text{WeightingFactor}(T_k)$ will be.

2. Text vectorization

Text vectorization is used to represent text as a real number vector that can be recognized by a computer. According to different granularities, text feature representation can be divided into several levels: word, sentence or text. Most classification algorithms are only applicable to discrete numerical types, so the text is quantized in space, that is, a text is represented by multidimensional feature vectors. For example, take the following two documents:

d1(A, B, C, D, E, F, G)

d2(C, E, F, G, A, B)

d1 and d2 are expressed as vector models, and the values of each dimension in parentheses are expressed by document characteristics. Then, we convert the feature words into numerical expressions, which are expressed by probability. This is illustrated by this example. Table 3 shows the number of feature words in the document.

Table 3. Number of feature words in the document.

Word	Document 1	Document 2	Document 3
A	2	1	3
B	2	2	4
C	2	1	3
D	1	0	1
E	4	3	7
F	2	1	3
G	2	2	4
Total words	15	10	25

The next step is to calculate the probability and standardize it. To prevent text features from being controlled by feature words with large values, the values are mapped between -1 and 1 . The calculated values are shown in Table 4. Table 5 shows the reverse frequency of feature words in documents. Finally, the resulting text vector table is shown in Table 6.

Table 4. Frequency of feature words in documents.

Word	D1	D2
A	0.08	0.04
B	0.08	0.08
C	0.08	0.04
D	0.04	0.00
E	0.16	0.12
F	0.08	0.04
G	0.08	0.08

Table 5. Reverse frequency of feature words in documents.

Word	Reverse Frequency
A	0.4
B	0.4
C	0.4
D	1.1
E	0.4
F	0.4
G	0.4

Table 6. Representation probability value of feature words.

Word	d1	d2
A	0.032	0.016
B	0.032	0.032
C	0.032	0.016
D	0.044	0.000
E	0.064	0.048
F	0.032	0.016
G	0.032	0.032

Finally, d1 and d2 can be expressed as space vectors as follows: d1 = (0.032, 0.032, 0.032, 0.044, 0.064, 0.032, 0.032); d2 = (0.016, 0.032, 0.016, 0.000, 0.048, 0.016, 0.032). The dimension feature variables corresponding to the vector are: A,B,C,D,E,F,G.

3. Cosine similarity model

The basic principle of cosine similarity is to use the cosine value of the angle between two vectors in the vector space as a measure of the difference between two individuals. The closer the value is to 1, the closer the angle is to 0° , with the two vectors being more similar. The model measures the similarity between two variables in all directions (attributes). The cosine value between two vectors can be determined by using the following formula.

$$a \times b = \|a\| \|b\| \cos(\theta) \quad (12)$$

Given two attribute vectors, A and B , the similarity of other chords is given by the following formula:

$$\text{similarity} = \cos(\theta) = \frac{A \times B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (13)$$

In Formula (13), A_i and B_i represent the components of vectors A and B , respectively.

The cosine similarity calculation is a part of the collaborative filtering algorithm. The key problem with a collaborative filtering algorithm-based user is how to calculate the similarity between users. There are three common algorithms for computing user similarity: cosine similarity, the Pearson coefficient and adjusted cosine similarity. These three methods are calculated based on the data structure of the user item matrix. The latter two methods are described below. The Pearson coefficient similarity method measures the similarity of two users through the Pearson correlation coefficient. The core idea of this method is to first find the item set that two users score together and then calculate the correlation coefficient of these two vectors. The formula is as follows:

$$\text{sim}(i, j) = \text{corr}_{i,j} = \frac{\sum_{u \in U} (R_{u,i} - R_i)(R_{u,j} - R_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - R_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - R_j)^2}} \quad (14)$$

In the formula, $R_{u,i}$ represents the rating of item i given by user u . The core idea of adjusting cosine similarity is to subtract the average user score vector from the vector in cosine similarity. The cosine of the included angle is calculated to correct the difference of the scoring scales of different users.

In the case of sparse data, these methods all have certain problems: For items not evaluated by the user, the score of the adjusted cosine similarity method is 0, and the set of items scored jointly by users in the Pearson coefficient may be very small.

To solve the sparsity problem in terms of user data. This paper proposes the following methods to eliminate this sparsity:

- (1) If you score from top to bottom, the sub item score and the parent item score will be calculated according to certain rules as the final score. Thus, the sparse matrix is filled into a dense matrix.
- (2) This method recommends user data according to the dense matrix.

Therefore, this paper proposes tourism text classification based on cosine similarity. The specific steps are as follows: (1) According to the keywords related to cultural tourism, this paper constructs a customized thesaurus with strong cultural tourism relevance; (2) word segmentation for each travel article is conducted; (3) then, the TF-IDF algorithm is used to give weight to the word segmentation and extract the first 15 words with high frequency; and (4) finally, the cosine similarity model is used to calculate the similarity value of the extracted tourism products.

4.1.2. Tourism Text Classification Based on Naive Bayes

The classification of tourism texts is mainly divided into two processes: the training process and the testing process. The main goal of the training process is to obtain a text classifier based on the existing training data. This paper proposes a tourism text classification based on Naive Bayes. The main steps are:

Step 1: The tourism text is preprocessed, tourism description text is automatically segmented and stop words are removed;

Step 2: Through word frequency analysis of the relevant cultural topic texts, the key words of cultural topics are manually selected, and the key vocabulary of cultural topics is constructed;

Step 3: The word bag model text vectorization method is used to represent the tourism description text;

Step 4: Feature selection is carried out via the information gain method;

Step 5: The TF-IDF method is used to assign values to lexical features, and the weights are changed according to the tourism keyword database;

Step 6: The naive Bayesian algorithm is used to classify tourism text. The formula is as follows:

$$P(d_i|c_j) = \frac{1 + \sum_{t=1}^{|D|} B_{it}P(c_j|d_i)}{2 + \sum_{t=1}^{|D|} P(c_j|d_i)} (j = 1, 2, \dots, |C|; t = 1, 2, \dots, n) \quad (15)$$

Step 7: The data in the test process are processed in the same way and filtered according to the characteristics of the training process. Finally, the trained naive Bayesian classifier is used to classify the text.

4.2. Experimental Results of Tourism Text Classification

This study collected 22,600 tourism texts for classification experiments, and compared them in terms of five aspects: service, location, facilities, health and cost performance. Two text classification methods were adopted: one is the naive Bayesian method and the other is the cosine similarity method. For the evaluation of the classification effects, we used the accuracy, precision, recall and F1-score in the classification model evaluation indicators to conduct experiments. The comparison results of the accuracy, precision, recall and F1-score of the two classification algorithms are shown in Figure 2.

Through the comprehensive analysis of the experimental results, the cosine similarity classification method has a significant advantage in the classification evaluation indicators, so it was more suitable for the tourism text collected in this paper.

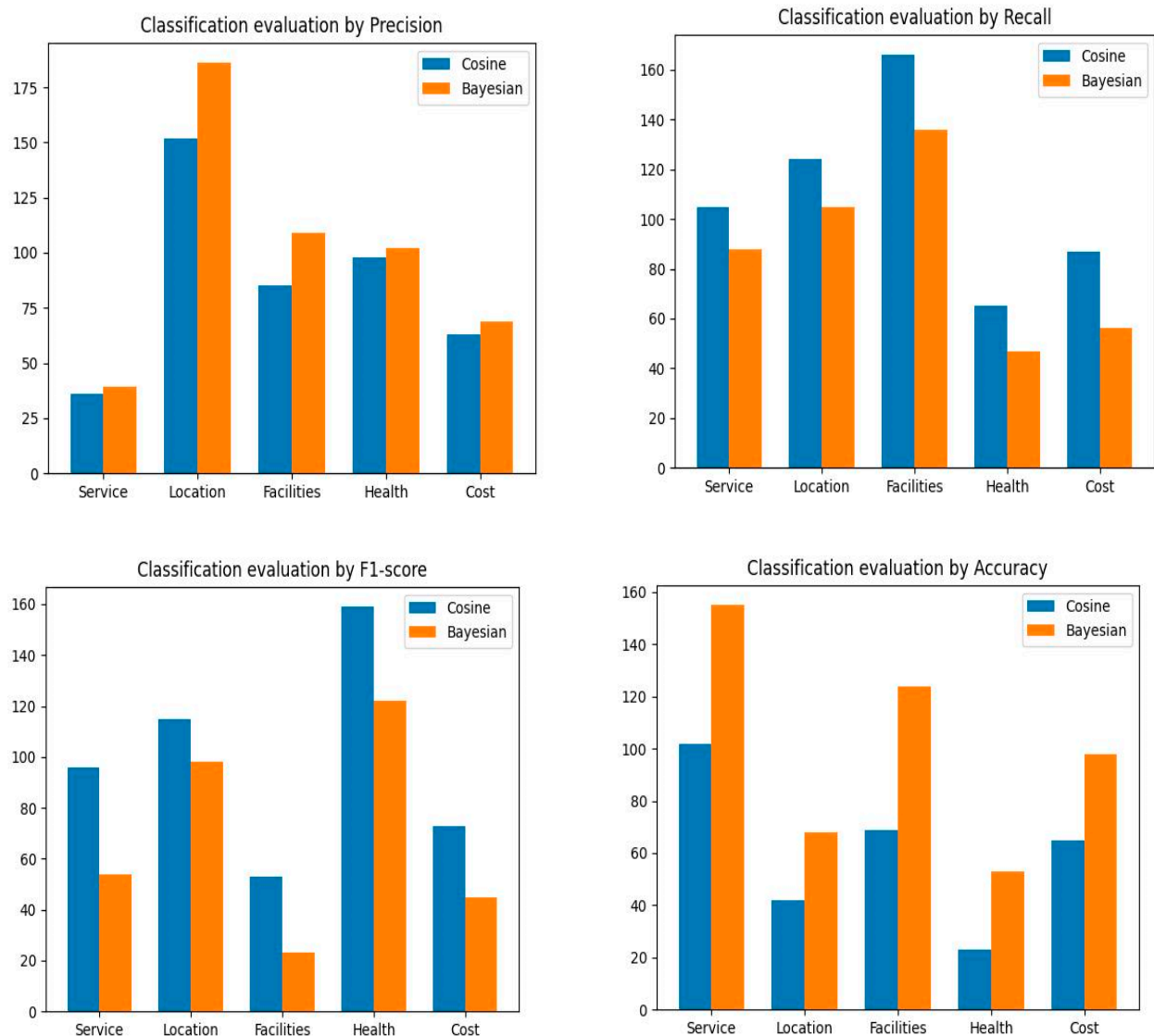


Figure 2. Accuracy, precision, recall and F1 score comparison results of two classification algorithms.

5. Analysis on the Popularity of Tourism Products

In order to obtain the popular words in tourism, it is necessary to conduct a popularity analysis of tourism products. This paper adopts two methods to build the popularity analysis model, including: one based on the principal component analysis method and another based on the sentiment analysis model. In order to verify the two popularity analysis models, Recall@K and Precision@K were selected as evaluation indicators to carry out popularity analysis comparative experiments.

5.1. Methods for Tourism Products Popularity Analysis

5.1.1. Construction of Multidimensional Tourism Product Popularity Analysis Model Based on Principal Component Analysis

Popularity is an important indicator for tourists to choose tourism related products, which directly reflects the popularity of products and passenger flow. According to the OTA and UGC data provided, examples of scenic spots, hotels, online popular scenic spots, home stays and special catering tourism products were extracted.

When analyzing the popularity of scenic spots, hotels and restaurants, it is necessary to identify the entity attributes contained in the opinion sentences. Then, through the attribute-based emotional analysis, the comprehensive emotional scores of scenic spots,

hotels and restaurants with respect to the five aspects of service, location, facilities, health and cost performance are calculated respectively. The user's attention to each evaluation dimension and the number of comments are used as weights to establish a multi-dimensional popularity analysis model. The next step is to calculate the total score of each scenic spot, hotel and catering. In this regard, this paper provides a multi-dimensional comprehensive popularity analysis of the destination.

1. Extract tourism products

Based on the collected tourism data, it is possible to extract specific tourism products and relevant useful information from scenic spot reviews, catering reviews and hotel reviews. First, define the names of tourism products in the comments as scenic spot names, catering names and hotel names.

Taking hotel reviews as an example, we extract product names to facilitate data consolidation. The specific steps are as follows:

Step 1: Group by product name through 'groupby' function;

Step 2: Use the apply function to perform the lambda expression for a multi column operation;

Step 3: Delete the original product name column.

Because the groupby function causes index disorder, we can use the reset_index function to reset the index. In Python, the groupby function can perform data grouping and intra-group operations after grouping to obtain a groupby object without any operations.

We assign the product names one by one. Through the vlookup function in Excel, the newly listed products are sorted out. Next, we will do the same for scenic area reviews and catering reviews.

2. Construction of tourism product popularity analysis model

For multi-dimensional popularity analysis, we chose to analyze tourism products from three perspectives. We weighted the product popularity from three dimensions: emotional score, frequency of occurrence and effectiveness of comments. SnowNLP is a class library written in Python which is a natural language processing library for Chinese. It has complete functions, including Chinese word segmentation, word annotation, emotion analysis and text classification, etc.

The three dimensions are described below. (1) The emotional analysis in this paper involves calculating the emotional score of comments by building a model through SnowNLP. The SnowNLP emotion score range is 0–1. The closer to 1, the more positive the emotion is. On the contrary, the closer to 0, the more negative the emotion is. (2) The frequency is the number of occurrences of each product. The more occurrences of a product, the higher the popularity of the product. (3) The validity is the number of words in each comment. For a product, the number of words in a comment will also affect the popularity.

Since the magnitudes of the three dimensions are different, we used standardization to map the validity and frequency to 0–1. Since the weight proportion of the three dimensions will affect the analysis result in terms of popularity, we used principal component analysis to determine the weight of the three dimensions. The results are shown in Table 7.

Table 7. Three-dimensional weight table in annual popularity analysis.

Year	Frequency (W1)	Sentiment (W2)	Length (W3)
2018	0.468	0.31	0.22
2019	0.414	0.31	0.277
2020	0.421	0.315	0.263
2021	0.431	0.31	0.259
Average	0.4335	0.31125	0.25475

According to the experimental results in Table 7, the popularity formula is as follows:

$$P(W) = W_1 \times Frequency + W_2 \times Sentiments + W_3 \times Length \quad (16)$$

This paper proposes a multi-dimensional tourism product popularity analysis model based on principal component analysis. The number of tourism products collected in the experiment reached 13,890. The experiment was primarily used to calculate the popularity value of tourism products in 2018, 2019, 2020 and 2021. The popularity values were standardized between 0–1 in the experiment.

5.1.2. Construction of Tourism Product Popularity Analysis Model Based on Emotion Analysis

1. Extracting tourism products based on ATF * PDF model and traversal method

There is no specific tourism product in the WeChat official account and travel guide, so a model needs to be established to extract the names of tourism products. The tourism strategy is a descriptive tourism text with an irregular structure, and its content is very complex and diverse. We have adopted two methods to extract tourism products from tourism strategies. The first involves using ATF * PDF model extraction. The second is extraction through traversal.

(1) Extracting tourism products based on ATF * PDF model

The steps of the method are: (1) carry out Chinese word segmentation and the removal of stop words for tourism strategies; (2) through the word frequency analysis of the tourism related words provided in the topic, select the key words of tourism topics, and build a key vocabulary of tourism topics; (3) use VSM to represent scenic spot description text; (4) carry out feature selection by the information gain method; (5) assign the lexical features with the ATF * PDF method as the weight, and change the weight according to the key vocabulary of cultural topics; (6) finally, extract the tourism products.

(2) Extract tourism products by traversal method

The steps of the method are: (1) Create a custom dictionary with two parts: tourism products extracted from scenic spots and hotels data above; search the internet to obtain tourism products such as scenic spots, hotels and popular online attractions; (2) traverse travel strategies through a user-defined dictionary to extract tourism products matching the dictionary; (3) extract product names from scenic spot reviews, hotel reviews and tourism products to be duplicated; (4) finally, map the product names extracted from scenic spot reviews, hotel reviews and tourism products to the corpus ID.

2. Emotional analysis

The primary purposes behind text emotion analysis is to analyze subjective text with emotional factors, excavate the emotional tendency contained in it and divide the emotional attitudes. The process of text sentiment analysis includes data preprocessing, feature extraction, classifiers and the output of sentiment categories. Two methods of emotion analysis are introduced below.

(1) Emotion analysis method based on emotion dictionary

The method based on emotion dictionary divides the emotion polarity in different granularities according to the emotion polarity of emotion words provided by different emotion dictionaries. The process of this method is: to input the text first and then preprocess the data (including denoising, removing invalid characters, etc.). Then, the word segmentation operation is carried out. Words of different types and degrees in emotional dictionaries are trained. Finally, the emotion type is output according to the emotion judgment rules.

(2) Emotion analysis method based on machine learning

The main idea behind the emotion analysis method based on machine learning is to extract features from a large number of labeled or unlabeled corpora using statistical machine learning algorithms and finally output the results of emotion analysis. The emotion classification methods based on machine learning can be divided into three categories: supervised, semi-supervised and unsupervised methods. Common supervised methods include KNN, naive Bayes, SVM and decision tree.

3. Construction of popularity analysis model

The popularity of tourism products reflects the popularity of a certain tourism product among people at a certain time. When defining the heat evaluation indicators, it is necessary to comprehensively consider the time period, the number of relevant comments and their emotional orientation. By analyzing the evaluation indicators of popularity, this paper provides a popularity analysis model of tourism products, as is shown by the following formula:

$$P(w) = -\log_2 \frac{1}{\alpha + 1} + 0.3 \times \frac{\alpha}{\lambda + \alpha} + 0.7\beta \quad (17)$$

In the above formula, α represents the frequency at which tourism products are mentioned, λ represents the maximum time difference of a product review and β represents the emotional score of tourism products.

The next step is to normalize the heat value to 0–1, as shown in the following formula:

$$w_i = \frac{m_i}{\sum_{i=1}^n m_i} \times 100\% \quad (18)$$

In the formula, m_i represents the popularity value of the i th tourism product.

5.2. Experimental Results of Tourism Products Popularity Analysis

This paper proposes a tourism product popularity analysis model based on emotion analysis, as shown in Formula (17). Experiments were carried out to verify the effectiveness of the heat analysis model. A total of 21,900 tourism products related to emotional analysis were collected. The experiment calculated the popularity value of tourism products in 2018, 2019, 2020 and 2021, respectively. The popularity values were normalized according to Formula (18).

In order to validate the tourism product popularity analysis model, we selected Recall@K and Precision@K as the evaluation indicators to analyze the generated tourism products in terms of popularity. The performances of the models based on principal component analysis (PCA) and the model based on emotion analysis (EA) were compared for the collected tourism product data in 2018, 2019, 2020 and 2021. The experimental results are shown below.

From the experimental results in Figure 3, it can be seen that the recall rates of both models increase significantly with the increase in K value. The recall rate of the PCA model is 2% higher than that of the EA model, which indicates that PCA has obvious advantages in the recall rate as an evaluation indicator. As can be seen from the experimental results in Figure 4, the precision of both models tends to decrease with the increase in K value. The precision of the EA model is 1.6% higher than that of the PCA model, indicating that EA has an obvious advantage in precision as an evaluation indicator. Therefore, when we conduct the popularity analysis of tourism products, we can synthesize the two models and average the popularity values obtained from the two models, so as to obtain a more reasonable tourism product popularity value.

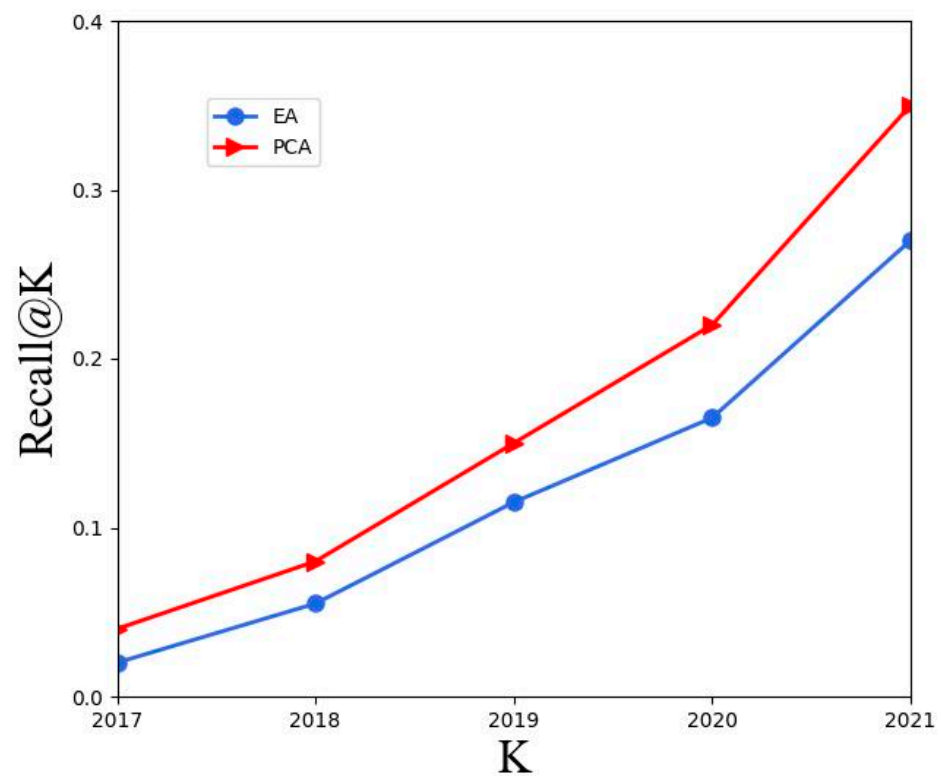


Figure 3. Experimental results of Recall@K on analysis popularity of tourism product model.

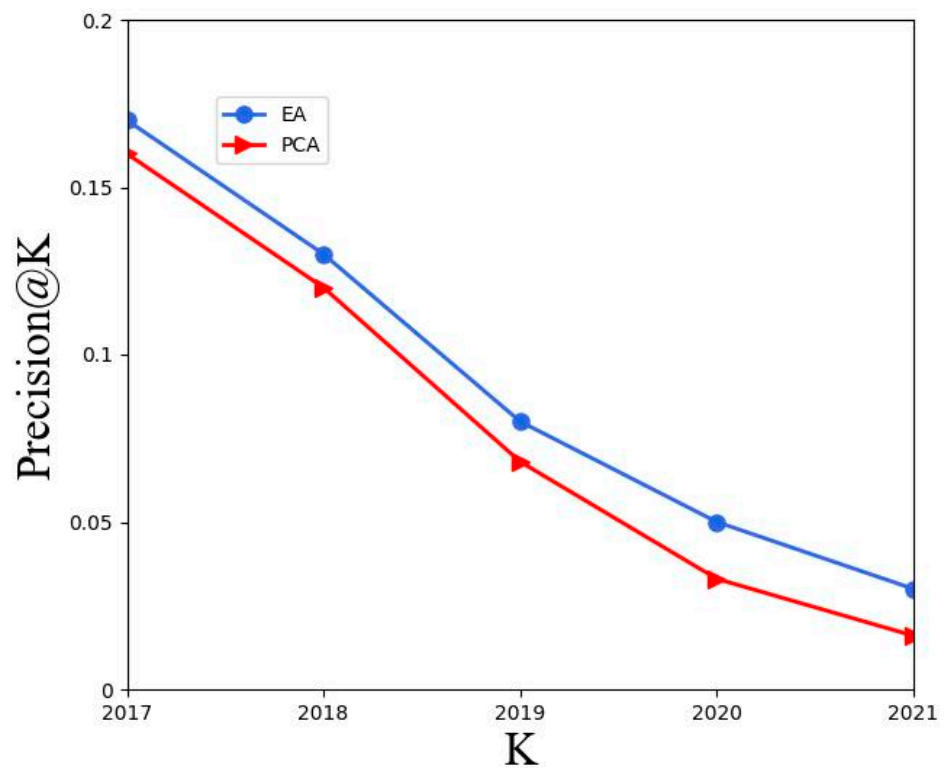


Figure 4. Experimental results of Precision@K on analysis popularity of tourism product model.

6. Building and Analysis of Tourism Knowledge Graph

A knowledge graph is a semantic network that reveals the relationships between entities. Knowledge graphs can harness, preserve and organize massive amounts of data, bringing internet information representations closer to human forms of cognition. The

process of building a knowledge graph starts with the pre-processing of raw data. Data sources may be structured, unstructured or semi-structured. Through a series of automatic or semi-automatic technical means, knowledge elements are extracted from the original data. This generates a mass of entity relationships (entity–relationship–entity) or entity attributes (entity–attribute–value). The extracted knowledge elements are fused, with this primarily including entity alignment and attribute alignment. The process of building a knowledge graph can lead to the discovery of new knowledge based on certain inference rules. The knowledge is stored in the pattern and data layers of the knowledge base after quality evaluation. The process of constructing the tourism knowledge graph proposed in this paper is shown in Figure 5.

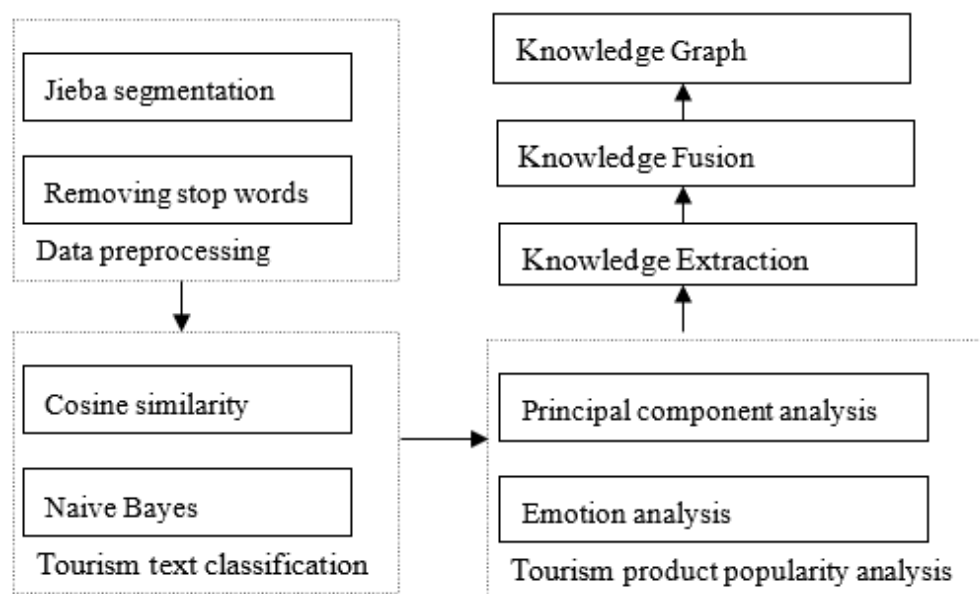


Figure 5. Construction process for the tourism knowledge graph.

6.1. Methods for Building Tourism Knowledge Graphs

6.1.1. Knowledge Extraction Based on BiLSTM-CRF

We needed to build an unstructured dataset before building a knowledge graph. We stored multi-dimensional tourism products into a triplet model. Knowledge extraction is the most basic process involved in the construction of a tourism knowledge graph. According to the results of the popularity analysis of tourism products, the knowledge of product types, product names and product association types was extracted.

The following problems are encountered after entity extraction: (1) The entities extracted from each tourism text are the same product, but the names are different, which may lead to ambiguity, and (2) the different quantities of products extracted from each article will lead to a different “length” in terms of the extracted products, which will lead to null filling. In order to solve the above problems, we use CRF method for named entity recognition. The CRF algorithm is briefly introduced below.

CRF is a kind of sequential annotation model algorithm. The BiLSTM model combines text context information [28]. Its principle is to infer the corresponding state sequence according to the given observation sequence and use the label relationship before and after the adjacent to obtain the current optimal label. Therefore, this paper overlays a layer of CRF after the output of BiLSTM to mark the category of named entities.

The CRF model is a sequence annotation on the sentence level. In order to effectively predict the relationship between tags, the CRF model uses the dependency information between tags at the sentence level. In this paper, the state transition matrix Q is defined as a parameter adjoint model to train together. Q_{ij} represents the transfer probability value,

that is, the probability value of transferring from the i th label to the j th label. y is the tag sequence to be predicted, and the formula is as follows:

$$y = (y_1, y_2, y_3, \dots, y_n) \quad (19)$$

The prediction probability of the model for sequence y is determined by the word feature vector output by the attention level and the parameter matrix of CRF. Then the sum formula of the probabilities of each location is:

$$S(x, y) = \sum_{i=1}^n c_{i, y_i} + \sum_{i=1}^{n+1} Q_{y_{i-1}-y_i} \quad (20)$$

Therefore, this paper selects the BiLSTM-CRF model for entity recognition. The cosine similarity method proposed above is used to predict the relationship between entities. Finally, the triple of “scenic spot entity, attribute, attribute value” required for research is extracted. The core ideas behind this are: (1) BiLSTM is used to recognize and extract the global features of text sequence data; (2) the output of the network layer is connected to the CRF layer for learning the labeling rules between tags; (3) and the scores of each word calculated and the best label outputted.

6.1.2. Knowledge Fusion Based on BERT Model

After information extraction, entities, attributes and attribute values were extracted from massive heterogeneous texts. A triad of knowledge for the knowledge graph was preliminarily formed. However, the quality of these triads was not high enough, and there was a significant amount of redundant or even wrong information. Therefore, it was necessary to correct and fuse this information through knowledge fusion to ensure the reliability and integrity of the constructed knowledge graph.

The idea of knowledge fusion is to align triples in pairs, compare their entities, relationships/attributes and calculate their similarity. When entities and relationships/attributes are similar, alignment is required. The data source confidence of two triples is calculated. The entity, relationship/attribute of a low confidence triplet is replaced by the corresponding term of another triplet.

We first use the BERT model to represent the word features of self-monitoring learning on a large number of corpora and learn a significant amount of priori syntax and word meaning information from tourism texts. According to tourism entity recognition and relationship extraction tasks, small adjustments are made through BERT [29]. Finally, this method obtains the features related to the tasks in the tourism field.

Several word vector models are used to encode entities: one-pot, word vector trained by the word2vec toolkit and word vector trained by the BERT model based on character training. After the word vector represents the entity, the cosine similarity proposed previously is used to calculate the similarity between the two entities, as shown in the Formula (13).

When the similarity is greater than the threshold value, it is considered that two entities (relationships/attributes) are similar. According to the parameter adjustment experiment, the threshold of entity pair is set to 0.86, and the relationship/attribute similarity threshold is set to 0.91.

6.2. Results for Building a Tourism Knowledge Graph

6.2.1. Results of Named Entity Recognition Experiments

The experimental data in this paper are from the triplet data of the Chinese tourism knowledge graph constructed in the previous section. We separately extracted the triple data containing the entity profile from all the triple data as the final experimental data. After data pre-processing, the triple set, entity set, relationship set and their corresponding data dictionaries were established. Both the entity recognition model and entity relationship extraction model in this paper use the pre training language model BERT on the Chinese tourism corpus. The main parameter settings of the model are shown in Table 8.

Table 8. The main parameter settings.

Parameter Name	Entity Recognition	Relationship Extraction
Epoch	12	20
Max length	756	523
Batch_size	18	56
Learning rate	4×10^{-8}	4×10^{-6}

The super parameters of the experiment are shown in Table 9. The parameters represent the best parameters after combining experimental conditions and training effects, including network structure parameters and experimental super parameters.

Table 9. The super parameters of experiment.

Parameter Name	Parameter Value
Hide Layer Dimension of BERT	986
Batch size	38
Dropout	0.8
Number of BiLSTM layers	2
Learning rate	0.0002
Hide layer dimension of BiLSTM	562
Iterations	68

The model was applied to three groups of experiments on the tourism text corpus. We used the natural language processing tools LPT and Hanlp to label tourism data in sequence. In order to improve the effect of entity recognition, LPT and Hanlp were each integrated into BiLSTM-CRF for adjustment. The purpose of this experiment was to verify that the BiLSTM-LPT and BiLSTM-Hanlp models have indeed improved the efficiency of entity recognition. In order to better verify the recognition effect, the test set was primarily used for named entity recognition. We used the precision and recall general evaluation indexes to judge the performance of the model in the experiment.

$$precision = \frac{TP}{TP + FP} \quad (21)$$

$$recall = \frac{TP}{TP + FN} \quad (22)$$

In the above formula, TP represents the number of correctly identified entities or relationships in the test set; FP represents the number of entities or relationships identified as errors in the test set; and FN represents the number of unrecognized entities or relationships in the test set.

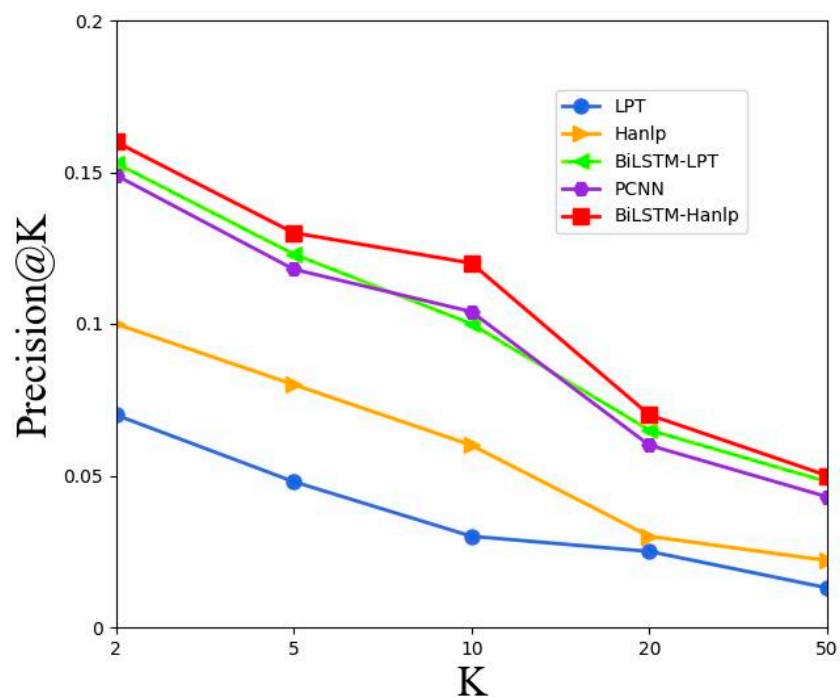
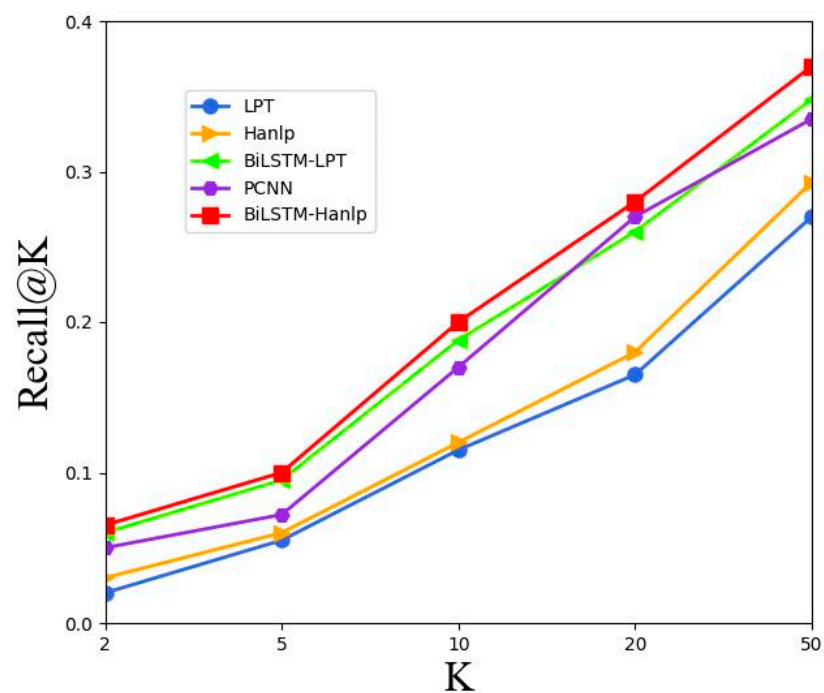
The deep learning baseline model selected in this paper is PCNN. The model separates the sentence into three segments by positioning the head entity and the tail entity. On this basis, the CNN convolution is introduced, and the maximum pooling is performed on each segment. The results of this comparative experiment are shown in Tables 10 and 11. The experimental results are shown in Figures 6 and 7.

Table 10. Experimental results of Precision@K.

Model	K = 2	K = 5	K = 10	K = 20	K = 50
LPT	0.07	0.048	0.03	0.025	0.013
Hanlp	0.1	0.08	0.06	0.03	0.022
BiLSTM-LPT	0.153	0.123	0.1	0.065	0.048
PCNN	0.149	0.118	0.104	0.06	0.043
BiLSTM-Hanlp	0.16	0.13	0.12	0.07	0.05

Table 11. Experimental results of Recall@K.

Model	K = 2	K = 5	K = 10	K = 20	K = 50
LPT	0.02	0.055	0.115	0.165	0.27
Hanlp	0.03	0.06	0.12	0.18	0.293
BiLSTM-LPT	0.06	0.095	0.188	0.26	0.348
PCNN	0.05	0.072	0.17	0.27	0.335
BiLSTM-Hanlp	0.065	0.1	0.2	0.28	0.37

**Figure 6.** Named entity recognition results based on Precision@K.**Figure 7.** Named entity recognition results based on Recall@K.

The experimental results show that two natural language processing tools, LPT and Hanlp, were successfully integrated into the model built in this paper. Therefore, the identification precision of the BiLSTM-LPT model is significantly higher than that of the LPT model. Compared with the model with only Hanlp added, the precision of the BiLSTM-Hanlp model increased by 15 percentage points. Compared with the PCNN model of deep learning, the BiLSTM-Hanlp model has obvious advantages in both precision and recall. The experiments show that the two natural language processing tools combined with the attention mechanism can improve the recognition effect of the model, and the precision and recall value are improved. Therefore, the entity recognition rate of a tourism knowledge graph based on BiLSTM+Hanlp+LPT model is significantly improved compared with the use of single natural language processing model.

6.2.2. Knowledge Graph Storage Based on Neo4j

Neo4j is a high-performance NOSQL graphic database, which stores structured data on the network rather than in tables. Neo4j can also be seen as a high-performance graph engine with all the features of a mature database. The data insertion and query operations in Neo4j are very convenient, regardless of the relationship between tables or databases. There are two basic data types in the drawing process: nodes and relationships. Nodes are connected through the relationships defined by relationships to form a relational network structure. Neo4j is characterized by diversification, beauty and powerful functions, so Neo4j was selected as the storage carrier of tourism knowledge in this paper.

The process of building a tourism knowledge graph presented in this paper is as follows. First, the corpus in the tourism text is screened to obtain the hot words of each tourist destination. Then, the processed data is transformed into a triple storage form for knowledge fusion. Then, the standardized data of the quality assessment is imported into the Neo4j graphic database. Therefore, 56,023 nodes and 86,500 attributes of entities are obtained. Finally, py2neo is used to construct and draw the local tourism knowledge graph.

7. Conclusions

In recent years, knowledge graphs, with their strong semantic organization abilities, have become the foundation of tourism knowledge-based organization and intelligent applications. The tourism knowledge graph in presented here connects the six elements of “food, accommodation, travel, shopping and entertainment” required by users in the form of triplets, facilitating the integration of complex tourism data.

This paper first analyzed the construction of tourism knowledge graphs at home and abroad, and founds that the existing tourism knowledge map has the problems of being time-consuming, laborious and having a single function. In order to build a high-quality tourism knowledge graph more efficiently, this paper has proposed a complete tourism knowledge graph system. The process of constructing the knowledge graph involves the following: In the data preprocessing stage, Jieba word segmentation and the removal of stop words are employed; in the text classification stage, after feature extraction and the vectorization of the text, the cosine similarity method is used to classify the tourism text; in the knowledge extraction stage, this paper proposes two methods to obtain popularity tourism words: One is a multi-dimensional tourism product popularity analysis model based on principal component analysis, and the other is a tourism product popularity analysis model based on emotional analysis; then, this paper proposes entity recognition based on the BiLSTM-CRF model, with the <tourism entity, attribute, attribute value > triplet being extracted; in the knowledge fusion stage, this paper adopts the pre training language model BERT and applies it to the Chinese tourism corpus. It integrates two natural language processing tools, LTP and Hanlp, and adopts the BiLSTM-LPT and BiLSTM-Hanlp models to improve the efficiency in terms of entity recognition. The experiments show that the proposed model can effectively improve the efficiency in terms of entity recognition. In the stage of knowledge graph generation, py2neo in Neo4j graphic database is used to build a high-quality tourism knowledge graph.

The limitations of this study include the following aspects: (1) Although this paper has constructed a relatively perfect tourism knowledge graphic, the influence of the time factor has been ignored in the construction process. That is, knowledge is time sensitive; (2) concerning the relationship extraction model proposed in this paper, there remain certain deficiencies in the application of the Chinese corpus. Therefore, attribute information of scenic spot entities should be added to improve the description of scenic spots.

Our future research includes the following. (1) The functions of the tourism knowledge graph system are relatively basic. In the future, certain functions can be added to the system, such as personalized recommendations by analyzing users' search, browsing and other records. (2) In the future, the expansion of existing data can reduce the consumption of manpower and time while automatically obtaining high-quality and high-volume training samples. (3) Additionally, in the future, more triplet information can be added to render the knowledge covered by the knowledge graph more comprehensive and substantial.

Author Contributions: Conceptualization, H.X.; methodology, H.X. and C.W.; formal analysis, G.F.; investigation, H.X. and G.K.; resources, H.X. and G.K.; writing—original draft preparation, H.X. and G.K.; writing—review and editing, H.X.; supervision, H.X. and G.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Funds of China (61272015) and 2022 Henan Province Key R&D and Promotion Projects (Science and Technology): “Research on Computer Monocular Visual Ranging Method and Application” (222102320342).

Data Availability Statement: Not applicable.

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

References

1. Paulheim, H.; Cimiano, P. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic. Web.* **2017**, *8*, 489–508. [\[CrossRef\]](#)
2. Jia, Z.H.; Gu, T.; Bin, C.; Chang, L.; Zhang, W.T.; Zhu, G.M. Tourism knowledge-graph feature learning for attraction recommendations. *CAAI Trans. Intell. Syst.* **2019**, *14*, 430–437.
3. Xiaoxue, L.; Xuesong, B.; Longhe, W.; Bingyuan, R.; Shuhan, L.; Lin, L. Review and Trend Analysis of Knowledge Graphs for crop pest and diseases. *IEEE Access* **2019**, *7*, 62251–62264. [\[CrossRef\]](#)
4. Alawad, M.; Gao, S.; Shekar, M.C.; Hasan, S.M.S.; Christian, J.B.; Wu, X.-C.; Durbin, E.B.; Doherty, J.; Stroup, A.; Coyle, L.; et al. Integration of Domain Knowledge using Medical Knowledge Graph Deep Learning for Cancer Phenotyping. *arXiv* **2021**, arXiv:2101.01337.
5. Tan, J.Y.; Qiu, Q.Q.; Guo, W.W.; Li, T.S. Research on the Construction of a Knowledge Graph and Knowledge Reasoning Model in the Field of Urban Traffic. *Sustainability* **2021**, *13*, 3191. [\[CrossRef\]](#)
6. Liu, W.; Liu, J.; Wu, M.; Abbas, S.; Hu, W.; Wei, B.; Zheng, Q. Representation learning over multiple knowledge graphs for knowledge graphs alignment. *Neurocomputing* **2018**, *320*, 12–24. [\[CrossRef\]](#)
7. Rizun, M. Knowledge Graph Application in Education: A Literature Review. *Acta Univ. Lodz. Folia Oeconomica* **2019**, *3*, 7–19. [\[CrossRef\]](#)
8. Shi, Y.; Gu, T.; Bin, C. Question and answer system of tourist attractions based on knowledge graph. *J. Guilin Univ. Electron. Technol.* **2018**, *38*, 296–302.
9. Zhang, H.; Khashabi, D.; Song, Y.; Roth, D. Transomcs: From linguistic graphs to commonsense knowledge. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, Yokohama, Japan, 7–15 January 2021; pp. 4004–4010.
10. Jiang, Z.; Chi, C.; Zhan, Y. Research on Medical Question Answering System Based on Knowledge Graph. *IEEE Access* **2021**, *9*, 21094–21101. [\[CrossRef\]](#)
11. Chiu, J.P.C.; Nichols, E. Named Entity Recognition with Bidirectional LSTM-CNNs. *arXiv* **2016**, arXiv:1511.08308v5. [\[CrossRef\]](#)
12. Duan, J.; Tan, Z.; Zhang, M.; Wang, H. New Word Detection Using Bi LSTM+CRF Model with Features. *IEICE Trans. Inf. Systems* **2020**, *103*, 2228–2232. [\[CrossRef\]](#)
13. Gasmi, K. Medical Text Classification based on an Optimized Machine Learning and External Semantic Resource. *J. Circuits Syst. Comput.* **2022**, *31*, 2250291. [\[CrossRef\]](#)
14. Gao, W.Q.; Huang, H. A gating context-aware text classification model with BERT and graph convolutional networks. *J. Intell. Fuzzy Syst.* **2021**, *40*, 4331–4343. [\[CrossRef\]](#)
15. Li, X.; Cui, M.; Li, J.; Bai, R.; Lu, Z.; Aickelin, U. A hybrid medical text classification framework: Integrating attentive rule construction and neural network. *Neurocomputing* **2021**, *443*, 345–355. [\[CrossRef\]](#)

16. Deng, J.; Cheng, L.; Wang, Z. Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification. *Comput. Speech Lang.* **2021**, *68*, 101182. [[CrossRef](#)]
17. Mughees, N.; Mohsin, S.A.; Mughees, A.; Mughees, A. Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting. *Expert Syst. Appl.* **2021**, *175*, 114844. [[CrossRef](#)]
18. Zhang, L.; Xu, C.; Gao, Y.; Han, Y.; Du, X.; Tian, Z. Improved Dota2 lineup recommendation model based on a bidirectional LSTM. *Tsinghua Sci. Technol.* **2020**, *25*, 712–720. [[CrossRef](#)]
19. Lv, L.; Wu, Z.; Zhang, J.; Zhang, L.; Tan, Z.; Tian, Z. A VMD and LSTM based hybrid model of load forecasting for power grid security. *IEEE Trans. Ind. Inform.* **2021**, *18*, 6474–6482. [[CrossRef](#)]
20. Xiong, B.; Meng, X.; Wang, R.; Wang, X.; Wang, Z. Combined Model for Short-term Wind Power Prediction Based on Deep Neural Network and Long Short-Term Memory. *J. Phys. Conf. Ser.* **2021**, *1757*, 012095. [[CrossRef](#)]
21. Lin, J.J.; Jian, Z.Q. Emotional Analysis of Cigarette Consumers Based on CNN and Bi LSTM Deep Learning Model. *J. Phys. Conf. Ser.* **2020**, *1651*, 012102. [[CrossRef](#)]
22. Kong, J.; Zhang, L.X.; Jiang, M.; Liu, T.S. Incorporating multi-level CNN and attention mechanism for Chinese clinical named entity recognition. *J. Biomed. Inform.* **2021**, *116*, 103737. [[CrossRef](#)]
23. Zhang, L.; Huang, Z.; Liu, W.; Guo, Z.; Zhang, Z. Weather radar echo prediction method based on convolution neural network and Long Short-Term memory networks for sustainable e-agriculture. *J. Clean. Prod.* **2021**, *298*, 126776. [[CrossRef](#)]
24. Lyu, C.; Chen, B.; Ren, Y.; Ji, D. Long short-term memory RNN for biomedical named entity recognition. *BMC Bioinform.* **2017**, *18*, 462. [[CrossRef](#)] [[PubMed](#)]
25. Song, J.; Tang, S.; Xiao, J.; Wu, F.; Zhang, Z. LSTM-in-LSTM for generating long descriptions of images. *Comput. Vis. Media* **2016**, *2*, 379–388. [[CrossRef](#)]
26. Xu, H.S.; Lv, Y.Q. Mining and Application of Tourism Online Review Text Based on Natural Language Processing and Text Classification Technology. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 9905114. [[CrossRef](#)]
27. Xu, H.S.; Sun, H.J. Application of Rough Concept Lattice Model in Construction of Ontology and Semantic Annotation in Semantic Web of Things. *Sci. Program.* **2022**, *2022*, 7207372. [[CrossRef](#)]
28. Bai, B.; Hou, X.; Shi, S. Named entity recognition method based on CRF and BI-LSTM. *J. Beijing Inf. Sci. Technol. Univ.* **2018**, *33*, 27–33.
29. Wu, Z.; Jiang, D.; Wang, J.; Zhang, X.; Du, H.; Pan, L.; Hsieh, C.-Y.; Cao, D.; Hou, T. Knowledge-based BERT: A method to extract molecular features like computational chemists. *Brief. Bioinform.* **2022**, *23*, bbac131. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.