Research on Tourism Destination Attraction Based on Deep Learning

Shanshan Li¹, Canhua Qiu¹ and Mingming Jiang²*

¹School of Economics and Management, Tongji University, Shanghai, China
²Bilibili Inc.

*stupidhod@gmail.com

Abstract. Mining and analyzing online travel reviews and travel information is playing an increasingly important role in the tourism industry. Accurately capturing the uniqueness and attractiveness of the tourist destinations recorded in the travel notes is the key to tourism analysis and application. The current way to obtain the attraction of tourism is easy to cause bias due to the use of simple statistical methods. This paper proposes a model based on deep learning, which uses Bert pre-training method, based on Transformer, and mines travel notes through Attention to find the attraction point. The model can understand the chapter-level semantics of travel notes based on the context, so much so that the extracted features are closer to the meaning of the text. It also exhibits good performance in generating unique labels of tourist destinations and similar tourism clusters. The experimental results are consistent with the facts, the validity of the model is also proved.

1. Introduction

The motivation of tourists to visit a city is often the uniqueness and attraction of the tourist destination. The scientific positioning and management of the attractiveness to the tourism image by managing teams can reduce the substitutability of the tourism destination and form a stable competitive advantage of the brand. With the development of Internet technology, the availability of extensive travel online reviews provides an unprecedented opportunity to analyse the emotions, preferences, feelings and opinions expressed by visitors. Therefore, it is important to determine the attractiveness and uniqueness of the image of a tourist destination by obtaining the text data of online travel platforms. It can help the tourists to choose the right place to travel and optimize the image of the tourist destination.

Currently, a large number of studies have been conducted through online analysis and application of online travel, and Serna [1] uses sentiment analysis techniques to analyse online hotel reviews and understand user preferences or requirements. Li Ping et al. used the website review data to extract the high-frequency feature words of the researched area and conduct semantic network analysis to obtain the image perception of five tourism communities in Beijing [2]. The measurement of the image of tourist destination is the subjective perception of the tourist towards the image of the tourist destination. It is also the basis for evaluating whether the tourist destination is attractive. The traditional method for measuring the image of a tourist destination is the surveying method by means of questionnaires or to carry out word frequency statistical analysis of the dimensional words to which the image of a destination belongs, etc. These methods all use simple statistical methods, which do not have a good understanding of the user's comments or travel notes and are likely to cause deviations in
results. From n-grams [3] to recurrent neural networks, technologies applied in the field of natural language processing are constantly evolving. Language modeling is the task of predicting the next word in a text given the preceding words, with practical applications in the NLP field. Transformer is a feature extractor that can train long text semantics and produce better results [4]. Google's BERT pre-training structure [5] solves the problem of long text dependence based on Transformer model. In view of this, this paper uses Bert pre-training model, based on Transformer structure to let the model learn the method of attraction mining with Mask's corpus. Then we use the trained model to extract the attraction of the travel notes. Finally, we show the attraction effect of the tourist attraction we extracted and the clustering results of similar tourist destinations.

2. Related work

2.1. Related technology application
Currently, word2vec[6] is one of the most widely used word embedding models for various applications related to text processing. But it still has some limitations. For example, it can't solve the problem of polysemous words. The learned word vector can't represent the global meaning. Moody proposed a model called lda2vec by mixing Dirichlet theme models and word embedding, which greatly improved the representation ability of standard word vectors [7]. The attention mechanism is used to learn to attribute the different weights of words to the overall meaning of the text, and is an effective technique for implementing long-term memory (LSTM). Li et al. proposed a two-way gated recursive unit neural network model (BiGRULA) by combining the topic model (lda2vec) and attention mechanism for analyzing hotel review data [8]. Dieng proposed TopicRNN, which integrates the advantages of RNN and potential topic models to achieve remote semantic dependence [9]. Zhou et al. [10] combined a convolutional neural network (CNN) with a long-term and short-term memory network (LSTM) to propose a C-LSTM model, which achieved good results on both the emotional and problematic classification tasks. The addition of LSTM, GRU models, and the Encoder-Decoder framework extends the application of RNN, but the sequence structure of RNN itself is difficult to perform large-scale parallel computing. CNN has strong parallel computing capabilities to solve text classification tasks, but CNN is more difficult to obtain long-distance features of text due to its structure. In the Transformer model [4], all calculations can be performed in parallel, thereby increasing the speed of training. Because of the existence of self-attention, there is direct interaction between any two frames, which can establish direct long-distance dependence.

2.2. Model structure
The model in this paper consists of four layers, which is shown in the figure1. The first layer is the input layer. Each [CLS]special symbol is added to the front of the vector representation, and the [SEP] special symbol is added at the end of the word embedding vector to mark and split the data. The vector is then position-coded with Position Embeddings and multiplied by the word vector of the mask corpus. The obtained characterization information enters the second layer of Transformer structure to obtain attractive points. These attractive points constitute a vector of 1x512 dimensions, multiplied by the vector of 512 cities, with 500 travel notes per city. Then softmax decide with which city's attraction it is best matched. The softmax parameter can represent the city's vector.
Figure 1. The model structure.

The bert pre-training method used in this paper only requires the Transformer model's Encoders architecture. This section will focus on the encoder architecture of the Transformer model. The encoder architecture is shown in the figure 2.

Figure 2. Encoder Architecture.

Each layer of the Encoder consists of two sublayers, a self-attention layer and a feed-forward neural network. At the bottom of the encoder input, the input word is first converted into a word embedding vector. The length of the input sequence is usually the length of the longest sentence of the training set, and each element is a 512-dimensional word vector. The bottom-level encoder takes the received sequence as input, then passes it through the self-attention processing, sends it to the forward network, and finally passes the input to the next encoder. After the input sequence is embedded in the word, the feature vector $Z$ is first obtained through the "Self-attention" module, and then passed to the fully connected layer of the feedforward neural network, and the output is passed to the next encoder. The formula for $Z$ is expressed as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}V\right)$$  \hspace{1cm} (1)

Multi-Head Attention is equivalent to the integration of $h$ different self-attentions and stitching together the final results. The multi-attention mechanism extends the ability of the model to focus on different locations, and multiple sets of randomly initialized query/key/value-matrix trained vectors can be mapped to different sub-expression spaces, giving multiple ways of expression to attention mechanism. The formula is expressed as:

$$\text{Multihead}(h) = \text{Concat}[\text{head}_1, \ldots, \text{head}_m]W^o$$  \hspace{1cm} (2)

where $\text{head}_i = \text{Attention}(Q,K,V)$ \hspace{1cm} where $Q, K, V = hW^0_i, hW^k_i, hW^v_i$

In addition, the location information of the word is added in the process of encoding the word vector, that is, the position code is introduced, so that the distance between two words is better expressed. The coding formula is as follows:

$$\text{PE}(\text{pos}, 2i) = \sin\left(\frac{\text{pos}}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \hspace{1cm} \text{PE}(\text{pos}, 2i + 1) = \sin\left(\frac{\text{pos}}{10000^{\frac{2i+1}{d_{\text{model}}}}}\right)$$  \hspace{1cm} (3)

pos: Location of words \hspace{1cm} I: dimensions of words

The add&norm layer is a combination of layer normalization [11] plus residual network in each sublayer. Add represents Residual Connection, which is to solve the problem of difficult training in multi-layer neural networks. By passing the information of the previous layer to the next layer, it can effectively focus on the difference. Norm represents Layer Normalization. Normalization of the layer's activation values speeds up the model's training process and allows it to converge faster. The related formula of feed-forward neural network is expressed as:

$$\text{FFN}(h) = \text{ReLU}(hW_1 + b_1)W_2 + b_2$$  \hspace{1cm} (4)

where $W_1 \in R^{d_{\text{model}} \times d_{\text{ff}}}$, $W_2 \in R^{d_{\text{ff}} \times d_{\text{model}}}$, $h \in R^{d_n \times d_{\text{model}}}$

2.3. Pre-training model bert
In October 2018, Google released a large-scale pre-training language model, BERT [11], the coding model structure of the two-way Transformer. In order to obtain the unique label of the tourist destination, the more accurate and efficient extraction of the information in the travel notes is the focus of this paper. This paper will use the BERT pre-training model. The BERT adopts a two-stage dual-language model, combining the dual advantages of the Transformer depth model and two-way information. The purpose of this pre-training is to learn the intrinsic logic of the text through two unsupervised tasks: Masked Language Modeling - predicting missing words by giving left and right contexts, and Next Sentence Prediction - Predict whether a sentence follows another sentence.

Before the data are put into the model, each word and special symbol of the input sentence need to be converted into a word embedding vector. The characterization information of the model input is obtained by three layers of Embedding. The first layer is Token Embeddings, and the Token Embeddings is a word vector. That is, the special character [CLS] and [SEP] are added the sentence word embedding vector, the special character [CLS] is used for the classification task, and the special character [SEP] is the sentence segmentation symbol. The final position coding Position Embeddings adds positional information to each word, allowing Transformer to perceive the positional relationship between different words.

In our model structure, softmax is the final output layer, and with the cross-entropy (CE) loss function. In this paper, we want to maximize the value of the \( p(y|x) \) probability in the model and the value of the cross entropy where \( y \) represents a tourist destination and \( x \) represents a travel note. By maximizing the value of \( p(y|x) \) probability, the difference between the travel information of this tourist destination and other travel destinations is found. In the Bert pre-training phase of our model, the input sequence of the fixed dimension is denoted as \( C \in \mathbb{R}^H \). During the fine-tuning, the only new parameter added is regarded as a classification layer \( W \in \mathbb{R}^{K \times H} \), where \( K \) is the number of classification labels, and the probability of the label \( P \in \mathbb{R}^K \) is calculated using the standard softmax, ie \( P = \text{softmax}(CW^T) \). Fine tune all parameters in BERT and \( W \) to maximize the log probability of the correct tag.

\[
p(y^{(i)} = j| x^{(i)}; \theta) = \frac{\exp(\theta^T_j x^{(i)})}{\sum_{i=1}^{K} \exp(\theta^T_i x^{(i)})} \tag{5}
\]

\[
l_{CE} = -\sum_{i=1}^{l_{CE}} \log(y^{(i)}) \tag{6}
\]

\( p(y^{(i)} = j| x^{(i)}; \theta) \) is the probability of categorizing \( x \) into category \( j \)

3. Data sources and processing methods

3.1. Data sources

Mafengwo is one of the relatively well-known travel websites in China and is the source of the data in this article. This article gathered the travel notes on the website of Mafengwo and obtained travels notes of 512 cities, each one with 500 to them, which means the total amount of the travel notes 256,000 travel notes. In addition to travel notes, this article also collects Chinese and English names for all tourist destinations in 512 cities.

3.2. Data processing

Because the neural network can only perform numerical calculations, this article will cut the numbers, English and each Chinese character in each travel note that we collected, and perform the word2vec operation to obtain a dense vector representation of the travel. In this paper, the BERT pre-training model is used. The length of the input sequence in the BERT pre-training model is 512 or less, including the [CLS] that the model needs to add before the input characterization information and [SEP] special symbols. Since the length of each travel note is long, in order to obtain the information of the data in the travel note as much as possible, this paper defines the vector length representation of the travel note as 512. After obtaining the word embedding vector representation of each travel note, if the length of a travel vector exceeds 510, the following information is discarded. If the length of a
travel vector is less than 510, the remaining length is replaced by the number "0". After that, the word embedding vector of each travel is multiplied by the [mask] vector with Chinese and English tourist destinations, and the information with Chinese and English tourist destinations in the travel notes is masked to obtain the travel words embedded vector after removing the names of Chinese and English tourist destinations.

4. Model results

4.1. Visualization of tourism destination attractions

In order to verify that our model can explore the uniqueness and attraction of the tourist destination in the travel notes, we randomly selected a travel note introducing Sichuan. We visualized the attention probability after each word in the travel note obtained by the model. And used the color scale in excel from red to gray to indicate probability of the last layer of the 12 headers from the minimum to the maximum. The red box of the 12 boxes behind the word means that the word is the attraction that the model notices. The deeper the red color, the more the word represents the uniqueness of the tourist destination. For reasons of space, we excerpted the two segments in this travel note to illustrate the attraction points of the model to Chengdu.

Because the word "Chengdu" has been processed by [mask], the probability of attention after "Chengdu" shown in the figure is zero, which is consistent with the purpose of our operation. Figure 3 shows that our model can choose words with emotions as attractiveness, such as “easy”, “gentle”, “leisure”, “trend”, “food”, “life”, and these words are also integral with our impression of the city. When it comes to Chengdu, people generally think of "panda" and "hot pot". These two words are also noticed by the model, as shown in the figure. Our model can learn the attraction of the tourist destination through a certain travel note, it has also verified the possibility that the model can learn the full attraction of a tourist destination by training all the travel notes of a tourist destination.

![Figure 3. Chengdu attraction points noticed by the model in a random travel note.](image)

4.2. Generation of all tourism destination attractions

Instead of a single character, the attraction of a tourist destination often consists of a word formed by more than one character. We get the travel notes of the Jieba word segmentation, and the Jieba participle belongs to the probabilistic language model segmentation, that is, to find the maximum probability in all the results obtained from the full segmentation. We add the attention information of the last layer of the multi-head to get the attention information on each character, and then re-add the attention information of each character in the dimension of words. Then we add attention information to the words of each of the 500 travel notes to get the probability of attraction of each word in this city. Then we sort them according to this weighted value W, the top word represents the attraction point of the tourist destination.

\[
C_{j-attention_k} = \sum_{i=0}^{11} P_{k-i-j}
\]  \hspace{1cm} (7)

\[
W_{attention} = \text{sum}(C_{i-attention}, C_{j-attention})
\]  \hspace{1cm} (8)
\[ P_{i->j} \]: attention information of the jth word on the i head
\[ C_{j->attention}^{k} \]: attention information for the jth word in k tokens
\[ i->j \]: index of the word in travel notes
\[ W_{attention} \]: attention information that makes up the word in all travel notes of a tourist destination

We selected two cities, namely Chengdu, Shanghai, and ranked the attraction points according to the attention scores of each city’s attraction, and get the 10 most attractive points with local characteristics. These attractive points are provided in Figure 4 and Figure 5. “Taikoo Li”, “Panda”, “hot pot”, “Delicious food”, “the land of abundance” and “Dujiang Dam” are the top six words that the model noticed in travel notes about Changdu. “Taikoo Li” is known as the most popular business street in Chengdu. Watching "Panda" and eating "hot pot" are two things that tourists almost always do when they come to Chengdu. Chengdu is also known as the “land of abundance”. The attractions of Shanghai include “Pearl of the Orient”, “Disney”, “Magic Capital” and “City God Temple”. These words also represent the characteristics and culture of Shanghai. “Oriental Pearl” is Shanghai's iconic cultural building. Shanghai “Disneyland” is the first Disney theme park in China. “Magic Capital” is synonymous with Shanghai, which is well known. Therefore, it is proved that the model excavates the attractive points of tourism destinations, including the subjective impressions of tourist destinations, special snacks, tourist attractions and humanistic scenes.

![Figure 4](image1.png)  ![Figure 5](image2.png)

5. Conclusion
In this paper, we propose a model based on deep learning that can explore the attraction and uniqueness of tourism. The model is characterized by the use of Google's latest BERT pre-training model, combined with the Transformer and Attention. The application of our model in online travel analysis shows that it can extract rich information from the text dataset, making it closer to the meaning of the article. The Attention scores the extracted tourism information to obtain the unique attraction point of the tourist destination. Using the parameters of the last layer of softmax as the embedding information of the urban tourism destination, the results obtained by the model are consistent with the facts and remarkably improves the traditional practice of conducting statistical analysis based on word frequency, making the attraction of tourist destinations more convincing and more in line with the essential characteristics of tourism destinations.

References
[1] He W, Tian X, Tao R, Zhang W, Yan G, and Akula V 2017 Application of social media analytics: a case of analyzing online hotel reviews. Online Information Review, 41(7), 921-935
[2] Li P, Chen T, Wang F and Wang Xinge 2017 Text Mining-based Perception of Urban Tourism Community Image: A Case Study of Beijing. Geographical Research 36 (6), 1106-1122.
[3] Kneser R and Ney H 1995. Improved backing-off for m-gram language modeling. In 1995 International Conference on Acoustics, Speech, and Signal Processing Vol. 1 pp 181-184
[4] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A. N and Polosukhin I 2017 Attention is all you need. In Advances in Neural Information Processing Systems pp 5998-6008
[5] Devlin J, Chang M W, Lee K and Toutanova K 2018 Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
[6] Mikolov T, Chen K, Corrado G and Dean J 2013 Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781

[7] Moody, C. E 2016 Mixing dirichlet topic models and word embeddings to make lda2vec. arXiv preprint arXiv:1605.02019

[8] Li Q, Li S, Hu J, Zhang S and Hu J 2018 Tourism Review Sentiment Classification Using a Bidirectional Recurrent Neural Network with an Attention Mechanism and Topic-Enriched Word Vectors. Sustainability, 10(9), 3313

[9] Dieng A B, Wang C, Gao J and Paisley J 2016 Topiernn: A recurrent neural network with long-range semantic dependency. arXiv preprint arXiv:1611.01702

[10] Zhou C, Sun C, Liu Z and Lau F 2015 A C-LSTM neural network for text classification. arXiv preprint arXiv:1511.08630

[11] Ba J L, Kiros J R and Hinton G E 2016 Layer normalization. arXiv preprint arXiv:1607.06450