Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 3, Issue 1: 24-33, DOI: 10.25236/AJCIS.030103

Cross-Attention Mechanism Recommendation Algorithm Based on LSTM

Mingming Qi[1,*], WenMing Ma[1], Rongjie Shan[1]

[1] School of Computer and Control Engineering, Yantai University, Yantai 264005, China
[*]Corresponding Author

Abstract. *With the development of economy, tourism is favored by more and more people, and location-based point-of-interest recommendation (POI) begins to emerge. This paper proposes the lstm-based recommendation algorithm for cross-attention mechanism (L-Attention). Embedding method was used to accurately learn the location sequence and location information of users, and feature information of location data was extracted through LSTM network. At the same time, the cross-attention mechanism is introduced to conduct dynamic modeling of personalized time check-in sequence, which can integrate the user's behavioral preference, location sequence information and the influence of sequential location on next-poi, so as to improve the accuracy of recommendation. This article uses Foursquare as an experimental data set. The results show that compared with the traditional recommendation algorithm which only considers the popularity of POI and the deep learning recommendation algorithm which only considers the timing, the accuracy of the proposed algorithm is greatly improved.*

Keywords: *Recommendation, points of interest, time series, LSTM, attention mechanism*

1. Introduction

With the rapid development of today's society, people's pace of life is gradually speeding up, and work pressure is also increasing. Therefore, tourism is favored by more and more people. There are numerous tourist destinations, and how to select POI[1-2] from numerous tourist destinations that are in line with users' personal interests has become an urgent problem for researchers, so location recommendation service emerges as The Times require. Location recommendation services refer users to places they might or would like to visit. The traditional recommendation algorithm is to recommend the most interesting travel places in the whole check-in point to the user, but it ignores the travel timing of the user. For example, the user

ISSN 2616-5775 Vol. 3, Issue 1: 24-33, DOI: 10.25236/AJCIS.030103

went to the zhongshan mausoleum in nanjing, the Chinese dinosaur park in changzhou and the taihu lake in wuxi three days ago, so it can be estimated that the user's place of play is in jiangsu. Considering the whereabouts of users on the fourth day, the greater probability is also in a tourist attraction in jiangsu. While the traditional recommendation algorithm may recommend the place in xinjiang or Tibet on the fourth day, which is obviously unreasonable. Therefore, timing problem is equally important for the recommendation of tourist sites. In view of the above unconsidered timing problems, this paper proposes a lstm-based POI recommendation algorithm model for the cross-attention mechanism. Feature information of location data was extracted through LSTM network, and the influence of travel timing on recommendation effect was dynamically solved by combining user-location cross-attention mechanism.

2. Related work

Liu et al [3] first generated a user-place score matrix based on the user's check-in records, then scored the places that the user had not visited by matrix decomposition , and Finally the locations of the top few predicted scores are recommended to users. This method makes up for the sparseness of the data to some extent, but it only considers the user's check-in records and ignores the context information of the place, and its recommendation effect is not ideal. Gao et al [4-5] considered the context information of the place when making the place recommendation to the user, and used the LDA topic model to fuse the relevant description information of the POI and the user's personal mood to generate a recommendation. This method combines the contextual information of the place, which has improved the accuracy to a certain extent. But how to better integrate the context information into the model has become a problem. Mikolo and Cheng et al [6-7] used the Word Embedding method to integrate the context information of the place into the model to solve the problem of POI recommendation, mainly by converting the context information into a low-dimensional space vector. The similarity between vectors is the next point of interest recommended by the user. With the development of neural networks, deep learning has become increasingly popular. Wang et al [8-9] used recurrent neural network RNN for modeling recommendations. RNN models can model sequence information well, and do not need to assume sequential dependencies relationship. Therefore, it is better than the traditional POI recommendation model in dealing with timing problems. Wang Li et al [10-12] proposed a POI personalized recommendation model based on LSTM. In this model, LSTM is essentially the same as standard RNN, but LSTM is richer internally, and can better handle complex social relationships and semantic information than RNN.

Compared with the above work, the next-poi studied in this paper recommended use LSTM network to extract the feature information of location data, and then generate location attention weight dynamically for the current temporal location through the cross-attention mechanism. After all, the importance of each location to the model is different, while the traditional deep learning model is based on the

ISSN 2616-5775 Vol. 3, Issue 1: 24-33, DOI: 10.25236/AJCIS.030103

weight of the overall model. In this way, the accuracy of POI personalized recommendation for users has been greatly improved.

3. Model description

This model is a POI recommendation algorithm model based on lstm's cross-attention mechanism. First, historical places visited by users, target place and users were embedded into the network model in a vectorized form through Embedding. Then, the user's historical location passes through the LSTM neural network layer in chronological order, while the target location and the user pass through the linear activation layer to obtain the output vector, target location vector and user vector of the user's continuous location respectively. Then the Attention model vector is obtained through a series of operations on the user vector and the sequential location vector. Finally, the user vector, the target location vector and the attention vector are stitched, and the stitched vector is passed through several linear hidden layers to obtain the final prediction result. The frame model is shown in Figure 1.

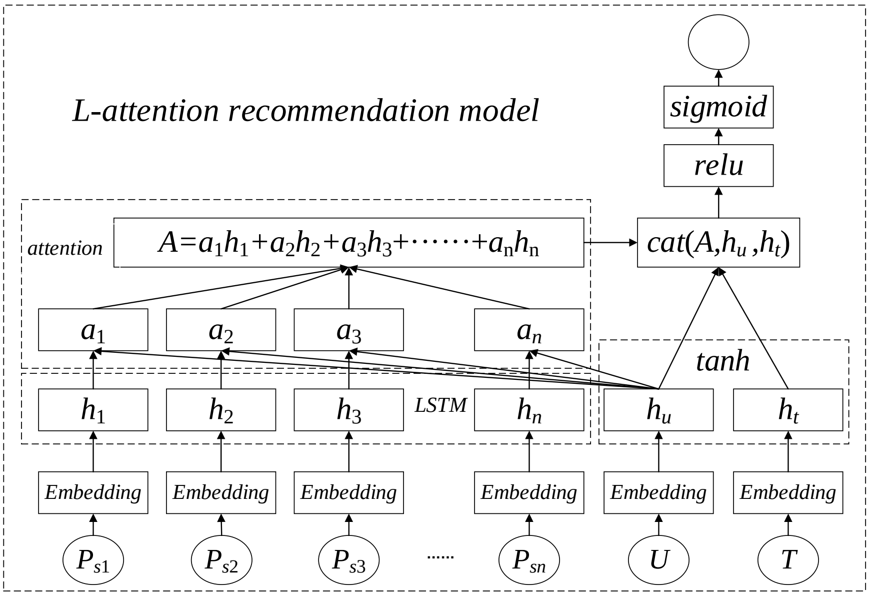

Fig. 1 Framework diagram of the model

The circular box represents the source input of the model, p_{s1} represents the first location visited by the user, and p_{sn} represents the NTH location visited by the user. U represents a specific user. T is the next destination for the user to visit after the location p_{sn}, also known as the predicted location. The rectangular box with Embedding is presented as the Embedding layer, and each source input enters the

ISSN 2616-5775 Vol. 3, Issue 1: 24-33, DOI: 10.25236/AJCIS.030103

network in the form of the Embedding word vector. The location of the user's historical behavior gets the output vector through LSTM, tanh means the hidden layer passing through is the linear layer, and the activation function is tanh. a_1, a_2, a_3 to a_n are the coefficients of LSTM output vectors h_1, h_2, h_3 to h_n under the cross-attention mechanism, and a_1h_1, a_2h_2 to a_nh_n are obtained by multiplying the coefficients and LSTM output vectors. A is obtained by multiplying each number times the corresponding position of the vector. Finally, the final prediction results were obtained by splicing A, h_u and h_t and feeding into the linear activation function layer. Sigmoid is the linear activation layer, the activation function is sigmoid, relu is the linear activation layer, and the activation function is relu.

3.1 Introduction to Attention

The Attention model literally means attention, that is, focusing on important points and ignoring unimportant points. Attention is generally divided into temporal attention and spatial attention. Temporal attention is mainly used in natural language processing. It is rarely involved in POI problems. The principle of using Attention is to calculate the degree of matching between the current input sequence and the output vector. The higher the match, the greater the weight of the place. The weight of matching degree calculated by Attention is only limited to the current sequence, instead of the weight of the whole as in the traditional neural network model. The specific meaning is explained as follows: the set of time and place sequences under user U is p_{s1}, p_{s2}, p_{s3} to p_{sn}, and the output through the LSTM layer is h_1, h_2, h_3 to h_n, and user U is output as h_u through the linear hidden layer. The specific operation is shown in formula (1) :

$$a_1 = h_1 \bullet h_u{}^{\mathrm{T}}$$

$$a_2 = h_2 \bullet h_u{}^{\mathrm{T}}$$

$$\ldots\ldots$$

$$a_n = h_n \bullet h_u{}^{\mathrm{T}}$$

$$a_1, a_2, a_3, \cdots a_n = soft\,max\left(a_1, a_2, a_3, \cdots a_n\right)$$

$$A = a_1h_1 + a_2h_2 + a_3h_3 + \cdots a_nh_n \tag{1}$$

h_1, h_2, h_3 to h_n and the transpose dot product of h_u to get a_1, a_2, a_3 to an respectively. a_1, a_2, a_3 to a_n after softmax are the coefficients of the output vectors h_1, h_2, h_3 to h_n of the LSTM layer. Finally, by multiplying and adding corresponding positions, the final vector A is consistent with the graph of the upper frame.

ISSN 2616-5775 Vol. 3, Issue 1: 24-33, DOI: 10.25236/AJCIS.030103

3.2 Introduction to LSTM

Long Term Memory network (LSTM) is a special type of RNN that can learn long-term dependent information. Since its inception, LSTM has been applied in many scenarios with great success. LSTM, like RNN, learns the sequence data through the chain form of repetitive neural network module. Meanwhile, to avoid the gradient explosion and gradient dispersion problem of RNN, LSTM increases the memory problem of long sequence by adding forgetting gate, input gate and output gate to the repetitive module of RNN.

Cell is the state maintained by the LSTM model and contains all the info rmation before the current time node. Input Gate is used to control the retent ion probability of the current Input information. Forget Gate is used to contro l the probability of retaining the information contained in the Cell according t o the current input and state. The Output Gate is used to control the Output probability of the information at the current time. Through the control of the three control gates, LSTM can continuously absorb input information, update its own state, and control the output. The forward propagation formula of L STM is as follows:

$$\text{Input Gate: } I_t = \sigma\left(W_i \cdot \left[h_{t-1}x_t\right] + b_i\right) \tag{2}$$

$$\text{Forget Gate: } F_t = \sigma\left(W_f \cdot \left[h_{t-1}x_t\right] + b_f\right) \tag{3}$$

$$\text{Output Gate: } O_t = \sigma\left(W_o \cdot \left[h_{t-1}x_t\right] + b_o\right) \tag{4}$$

$$\text{Kept state: } C_t^{'} = tanh\left(W_c \cdot \left[h_{t-1}x_t\right] + b_c\right) \tag{5}$$

$$\text{Cell State: } C_t = C_{t-1} * F_t + C_t^{'} * I_t \tag{6}$$

$$\text{Output: } H_t = C_t * O_t \tag{7}$$

Among them, σ is the sigmoid function, W is the weight in the neural network, b is the bias term, and tanh is the activation function of the output

4. Experimental results and analysis

4.1 Experimental data and environment

The data set used in this experiment is a long-term (about 10 months) boarding encrypted data set collected from Foursquare from April 12, 2012 to February 16, 2013 in New York City, which contains the user's check-in data and location Latitude and longitude and check-in time. Tested where each user has visited at least 100 locations.

Lab environment operating system: windows 7; CPU: Intel Core i5-6500; Main frequency: 3.2 GHz; RAM: 8G; Programming language: Python3.7; Deep learning framework: PyTorch1.2.0

4.2 Data preprocessing

Due to the encryption of experimental data, the data were processed as follows for the convenience of the experiment:

1) Unique encoding for 1083 users and 38,333 locations in the dataset

2) Eliminate sites that appear less than 10 times in the dataset

3) The places visited by users are arranged in ascending chronological order. Since this paper studies the Next-POI problem, the sequential arrangement is the focus. The arrangement results are shown in table 1.

Table 1 User place time sequence sorting table

User	Place
470	0,616,620,624,639,636…
979	1,669,1707,1712,868…
395	2816,786,8716,413,10530…
87	4,34,4349,5047,5307…
642	6,26,3101,3281,3694…
…	…

4) In the L-Attention model, the positive sample data of users are expressed as $\left\{\left[p_{s1},p_{s2},...p_{sn}\right],p_{sn+1}(T),U_i,1\right\}$. For example, for a user numbered 470, his positive sample data is $\left\{\left[0,616,624\right],639,470,1\right\}$, $\left\{\left[616,620,639\right],636,470,1\right\}$ etc, where 1 is the mark of positive sample. $\left[p_{s1},p_{s2},...p_{sn}\right]$ is the n locations that the user has been to continuously, $p_{sn+1}(T)$ is the next destination that the user will visit, and U_i is a specific user. From all positive samples, select a positive sample format data for each use r and put it into the test set to form the positive sample of the test set. The remaining positive sample format data is the positive sample of the training set.

5) In the L-Attention model, the negative sample data of users is expressed as $\left\{\left[p_{s1},p_{s2},...p_{sn}\right],X,U_i,0\right\}$. For example, for the user numbered 470, his negative sample data is $\left\{\left[616,620,...639\right],139,470,0\right\}$, $\left\{\left[0,616,624\right],998,470,0\right\}$ etc, where 0 is the mark of negative sample, $\left[p_{s1},p_{s2},...p_{sn}\right]$ is consistent with the positive sample, X is the negative sample of the user (any unrepeated place the user has not been to), U_i is consistent with the positive sample, and the positive and negative

ISSN 2616-5775 Vol. 3, Issue 1: 24-33, DOI: 10.25236/AJCIS.030103

samples of the training set are combined to form the training set and put into the model training. It should be noted that the proportion of positive and negative samples in the training set can be set as 1:1 in this paper.

6) generate positive samples and negative samples for the test set with a ratio of 1: n, and the data form is like the training set. Here, n is set artificially and set to 100 in this article. Both positive and negative samples of the test set constitute the test set and put it into the model test.

4.3 Benchmark experiments and evaluation standards

In order to verify the validity of the experiments in this article, the comparative experiments selected in this article are UP-RNN, POI-LSTM

The UP-RNN model is a POI recommendation model based on a recurrent neural network. This model uses RNN to output the user's point of interest vector, and then uses the output user's point of interest vector to stitch the user vector into the linear activation layer, and the output is used as the user to visit the place Probability.

The POI-LSTM model extends the long and short time memory neural network and uses the Embedding idea to vectualize user information and POI information and input it into the neural network. At the same time, LSTM is used to capture users' interest characteristics and interest change trends, and then various semantic information is fitted at different input layers to recommend the next interest location for users.

Evaluation criteria: As it is the prediction of the next point of interest, the evaluation criterion for this article is Precision and its formula is as follows:

$$Precision = \frac{\sum_{i=1}^{N} \sigma i}{N} \tag{8}$$

$$\sigma i \begin{cases} =1 & T \in sort_i^3 \\ =0 & T \notin sort_i^3 \end{cases} \tag{9}$$

Precision represents the proportion of users predicted to succeed in the total number of users. N is the total number of users, i represents a specific user ID, T represents positive sample of user. Each user in the test set has one positive sample and multiple negative samples. The number of negative samples for each user in this article is 100. The data of the positive and negative samples will eventually be output as the probability of the next place in the model. $sort_i^3$ represents the probability ordering of the positive and negative samples of the user in the test set, and the three probability sets with the highest probability. σi indicates that if the probability of the user's positive sample data is among the first three probabilities, then it is recorded as 1, otherwise it is recorded as 0.

ISSN 2616-5775 Vol. 3, Issue 1: 24-33, DOI: 10.25236/AJCIS.030103

4.4 Performance comparison

In order to verify the accuracy of the proposed algorithm, a comparative test was conducted between the proposed algorithm and the benchmark experimental algorithm. The accuracy of the recommended results is related to the sparsity degree of the data set, and the accuracy of the predicted results is also different when data sets with different sparsity degree are selected. Figure 2 shows the average precision values of 10 tests under different sparsity data sets selected by the three algorithms, in which we filter some training sets through the random function to control the sparsity of the training set. It can be seen that the accuracy of the algorithm in this paper is higher than that of the other two algorithms under different sparsity, and with the increase of data volume, the accuracy gap between the other two algorithms and the algorithm in this paper is obvious, which can reach about 10 percentage points. When the data set is large enough, the accuracy tends to be stable, so the algorithm in this paper has better performance in mining users' interest points and better recommendation effect.

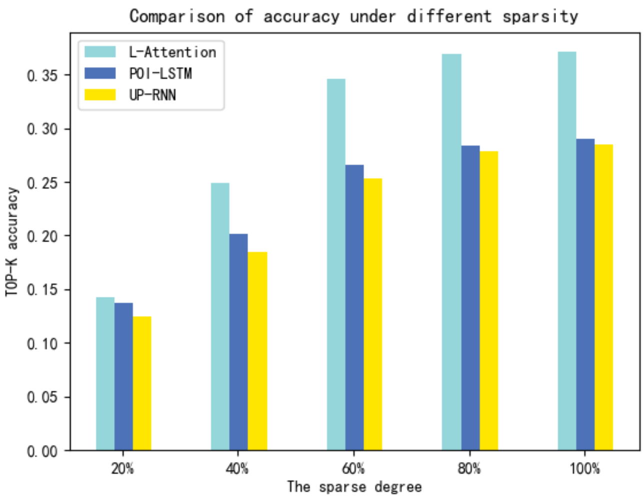

Fig. 2 Accuracy of each algorithm under different sparsity

Figure 3 is a comparison of the precision of sequence lengths at different locations under the same data set. The length of the sequence of places here is set artificially in this article. Through the experimental results can be seen that the experimental model accuracy under various site sequence length to be higher than the rest of the two algorithms, in particular, the optimal accuracy of the model is highest when the sequence length is 5, When the length is greater than 5, the accuracy of the model decreases, but it is also significantly higher than the other two algorithms, thus the algorithm has better performance than the other two kinds of algorithm.

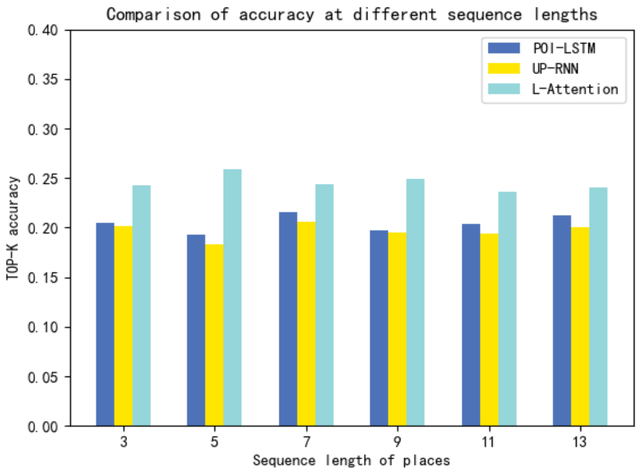

Fig. 3 Accuracy of each algorithm under different sequence lengths

5. Summary

This paper proposes a point of interest recommendation algorithm based on LSTM's cross-attention mechanism, which successfully combines LSTM and Attention and applies them to the location recommendation service scenario, thus solving the timing problems that traditional recommendation methods did not solve and the attention problems that were not considered under the deep recommendation algorithm Comparing the proposed L-Attention algorithm with other algorithms through real data sets, the results show that the recommendation model proposed in this paper is significantly better than UP-RNN and POI-LSTM in accuracy, thus improving the recommendation performance to some extent.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61602399).

References

[1] Berjani. B, & Strufe. T. (2011). A recommendation system for spots in location-based online social networks. In Proceedings of the 4th workshop on social network systems(p. 4). ACM.
[2] Ye. M, Yin. P. F, Lee. W. C, & Lee. D. L. (2011). Exploiting Geographical Influence for Collaborative Point-of-interest Recommendation International Acm Sigir Conference on Research & Development in Information Retrieval. ACM.

ISSN 2616-5775 Vol. 3, Issue 1: 24-33, DOI: 10.25236/AJCIS.030103

[3] Liu. B, & Xiong. H. (2013). Point-of-interest recommendation in location based social networks with topic and location awareness. In Proceedings of the 2013 SIAM International Conference on Data Mining (pp. 396-404). Society for Industrial and Applied Mathematics.
[4] Gao. H, Tang. J, Hu. X, & Liu. H. (2015). Content-aware point of interest recommendation on location-based social networks. In Twenty-Ninth AAAI Conference on Artificial Intelligence.
[5] Yin. H, Cui. B, Sun. Y, Hu. Z, & Chen. L. (2014). LCARS: A spatial item recommender system. ACM Transactions on Information Systems (TOIS), 32(3), 11.
[6] Mikolov. T, Chen. K, Corrado. G, & Dean. J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
[7] Cheng. C, Yang. H, Lyu. M. R, & King. I. (2013). Where you like to go next: Successive point-of-interest recommendation. In Twenty-Third international joint conference on Artificial Intelligence.
[8] Liu. Q, Wu. S, Wang. L, & Tan. T. (2016). Predicting the next location: A recurrent model with spatial and temporal contexts. In Thirtieth AAAI Conference on Artificial Intelligence.
[9] He. C, Liu. Y, Guo. Q, & Miao. C. (2019). Multi-Scale Quasi-RNN for Next Item Recommendation. arXiv preprint arXiv:1902.09849.
[10] Wang. Li, Zhang. Ye. (2018). LSTM-based POI personalized recommendation framework [J]. Computer System Applications, 27(12): 56-61.
[11] Zhao. P, Zhu. H, Liu. Y, Li. Z, Xu. J, & Sheng. V. S. (2018). Where to Go Next: A Spatio-temporal LSTM model for Next POI Recommendation. arXiv preprint arXiv:1806.06671.
[12] Zhou. Y, Huang. C, Hu. Q, Zhu. J, & Tang. Y. (2018). Personalized learning full-path recommendation model based on LSTM neural networks. Information Sciences, 444, 135-152.