Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

POI Recommendation for Social Relations based on WORD2VEC

Li Yang-yang, Wang Ya-jun*, Zhang Mi-yuan

Information and Communication Engineering, Liaoning University of Technology, Jinzhou, China

E-mail: *1179165495@qq.com

Abstract

Most of the traditional recommendation algorithm models are recommended based on the user's own historical preferences, although it can recommend POI for users to a certain extent. But in real life, people are more willing to ask their friends what they think when they have a difficult decision. Therefore, a word2vec-based social relationship point of interest recommendation model (W-SimTru) is proposed, which combines the similarity of friends based on cosine similarity with the friend trust recommendation algorithm based on TF-IDF to improve the model recommendation effect. In addition, before modeling the similarity of users, word2vec is used to process the user's historical check-in behavior to solve the problem of inaccurate recommendation due to sparse check-in data. Finally, experiments are carried out on three datasets of Los Angeles, Washington and NYC in Gowalla, and the experimental results show that the proposed W-SimTru recommendation algorithm outperforms the algorithms of the three comparative experiments.

Keywords: Points of interest, User similarity, Friend trust, TF-IDF, Word2vec

1. Introduction

Nowadays, with the vigorous development of mobile internet technology, the rapid increase in the number of POIs has caused more and more users to have "difficulty in choosing". How to help users to screen out the items and places that users may be interested in from the massive POI [1, 8-10] is the research direction of many researchers.Recommendation systems[2,3] have also flourished in this context.

The recommendation system uses the user's historical information to help the user decide what product to buy, and simulates the salesperson to help the customer complete the purchase process. Point-of-interest recommendation system [4,5] is a kind of

DOI: https://doi.org/10.36548/jaicn.2022.2.001
Received: 11.03.2022, received in revised form: 12.04.2022, accepted: 08.05.2022, published: 18.05.2022

recommendation system, which aims to help users explore new areas and discover new points of interest, thereby making location-based social networks more efficient. attractive.

Xie[6] designed a general potential friend recommendation framework by calculating the similarity based on interest features among users.Ye et al. [7] calculated the social similarity by calculating the ratio of common friends among multiple users. On this basis, it is combined with the traditional collaborative filtering method and named FCF. Wang et al. [8] proposed the idea of Embedding, quantified user information, friend relationship, POI information and input it into LSA to recommend the next POI for users.Zhang et al. [9] improved the traditional collaborative filtering algorithm by adding a regularization term for social relations. Also, combining with social information improves accuracy. Yang et al. [10] proposed a random walk scheme and combined sampling user check-in and social relations to recommend points of interest through the similarity of hypergraphs and hyperedges.

However, with the development of society, there are countless types of POIs, but the actual check-in data of users is very small, and the user check-in data is extremely sparse, resulting in inaccurate recommendations. In addition, most of the current recommendation algorithms only consider the preferences of users and POIs, and rarely consider users' social relationships, but in real life, users are more likely to visit POIs recommended by friends.

Therefore, in order to solve the above problems, this paper proposes a word2vec-based social relationship point of interest recommendation model. This algorithm first processes the user check-in data through word2vec [11,12], then calculates the similarity between users through the user cosine similarity, and calculates the trust between users' friends through TF-IDF, and recommends top-k POIs for users . Finally, experiments are carried out on three real datasets in Los Angeles, Washington and NYC, and compared with three traditional algorithms, the two evaluation indicators of precision and recall have been significantly improved.

2. User similarity recommendation algorithm based on word2vec

2.1 Word2vec

Word2vec is the process of converting a sparse word into a dense vector. The Word2vec model is mainly divided into the CBOW model and the Skip-gram model. This paper adopts the CBOW model, and its structure is as follows:

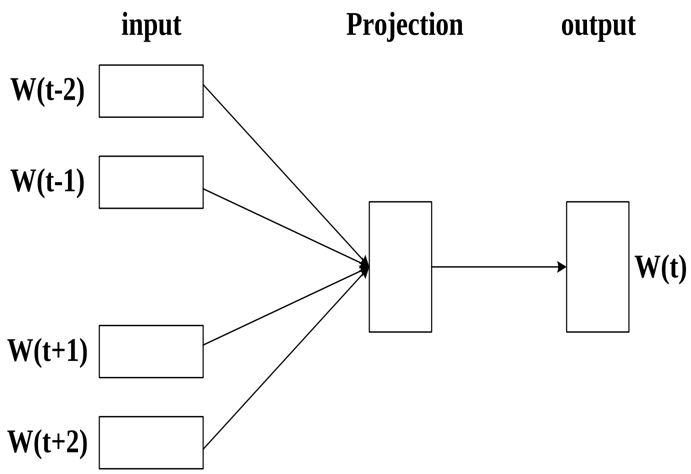


Figure 1. CBOW model

CBOW converts the concatenation layer of NNLM into summation, to reducing the running time of the model. The model predicts the central vocabulary according to the words around the central word W(t). The input information is several words before and after the target word, also known as the window. The input word vector is weighted and summed in the mapping layer and passed through the output layer to obtain the target. The predicted probability of the word. The input vector of the input layer is one-hot encoded $(\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_\mathrm{p})$, the output of the projection layer is the weighted average of the input vector, the formula is as follows:

$$\mathrm{h} = \frac{1}{C} W(\sum_{i=1}^{c} x_i) \tag{1}$$

Among them, W is the weight matrix of the projection input layer to the projection layer, C is the window size, is the word vector input by the input layer, and then the hidden layer is connected to the output layer through an N×V weight matrix W', U = h×W ', h is the output vector of the hidden layer, W' is the weight matrix connecting the hidden layer to the output layer.Finally, the value of the output layer is normalized by Softmax to obtain the output probability of the word.The CBOW model learns the weights through backpropagation. Gradient descent first defines the loss function. In the CBOW model, the loss function maximizes the conditional probability of the output word in a given input context, and then iterates through stochastic gradient descent to find the optimal parameters. Its loss function is:

$$\mathrm{Loss} = -\mathrm{logp}(w_i|\mathrm{context}(w_i)) \tag{2}$$

2.2 User similarity based on word2vec

2.2.1 Influence of user similarity on POI recommendation

Through research and analysis of previous articles, a well-known result can be drawn, that there are similarities in interests and hobbies between users. For example, if user 1 likes to eat cake and milk in the morning, and user 2 likes to eat cake, milk and eggs in the morning, then user 1 and user 2 are similar to a certain extent, so the eggs that user 2 likes to eat can be determined. Recommended to user 1.Therefore, considering the similarity between users in the recommendation algorithm will help to more accurately identify user preferences.

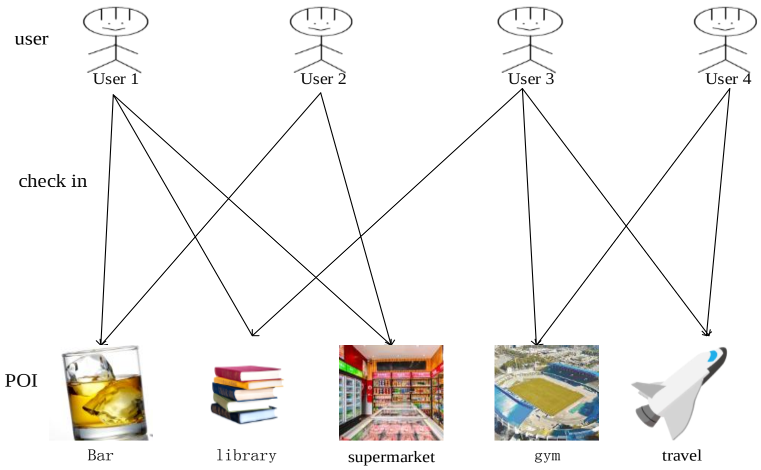


Figure 2. User sign-in diagram

Find other users who have the same or similar interests by analyzing the historical data of the target user's interests and preferences, and recommend POIs that have not been searched but may have interests for him, that is, whether there are the same or similar behaviors and interests is to see whether the two users have both Accessed (check-in) to certain items. If both parties have checked in on certain items, it is very likely that the two users are interested in each other's items that the other has visited, which proves that they are likely to have similar hobbies. Figure 2 takes 4 users and 5 points of interest as an example (in real life, there are thousands of users and points of interest), as shown in Figure 2, user 1 likes to go to bars, libraries and supermarkets, and user 2 likes to go to bars and supermarkets, then user 2 may like the library that user 1 likes to go to.

To sum up, some POIs checked in by the user in a certain segment are combined to form the user's historical check-in data set. Then, by calculating the similarity between the target user and neighboring users' historical POIs, and then according to the calculated weights, recommend POIs that may be of interest to the target user but have not been visited.

2.2.2 Cosine similarity

In this paper, the cosine similarity is used to calculate the similarity between two users. The formula is shown in (3):

$$\mathrm{H}(\mathrm{u},\mathrm{v}) = \frac{\vec{u}\cdot\vec{v}}{||\vec{u}||\times||\vec{v}||} \tag{3}$$

In the formula, $\vec{u}$ and $\vec{v}$ represent the user's embedding vector, $||\vec{u}||$ and $||\vec{v}||$ represent the modulus of the user's label vector.

Specifically, the user check-in data is first processed by the word2vec technology described in the previous section, and then the similarity between users is calculated by the cosine similarity, and then sorted according to the similarity as a candidate POI.

3. POI recommendation system based on social relationship

3.1 Friends Trust Recommendation Based on TF-IDF

Term Frequency-Inverse Document Term Frequency (TF-IDF), which is used to measure the importance of a word to a document, just like the trust of a user's friends to the user. For example, if user u1 has only one friend of user u2, then user u2 is more important to user u1. On the contrary, if user u3 has u4-u99 friends, the intimacy of user u4-u99 is low. The calculation formulas of TF and IDF are as follows:

TF word frequency:

$$TF_{u,v} = \frac{total_{u,v}}{total_u} \tag{4}$$

IDF inverse document word frequency:

$$IDF_v = \log(\frac{|U|}{total_v} + 1) \tag{5}$$

The $total_{u,v}$ value is 1, which $total_{u,v}$ is the number of all trusted friends of the user u, and $total_v$ is the number of friends of the friends trusted by the user. Calculate TF-IDF by TF and IDF:

$$TFIDF_{u,v} = TF_{u,v}IDF_v \qquad 6)$$

By calculating the weights of trusted friends, the POIs of friends are sorted to obtain candidate POIs.

3.2 Overall recommendation model based on social relations

This paper considers the user's social relationship (the user's hobbies are similar and the trust relationship between friends), Combining the similar relationship between users' POIs and the trust relationship between friends, it recommends POIs to users individually, and taps users' potential interests and hobbies. In addition, in order to reduce the negative impact of data sparseness, before calculating the similarity between users, word2vec processing is performed on the user's check-in data, and the high-dimensional sparse data is reduced in dimension. The overall recommendation framework is shown in Figure 3:

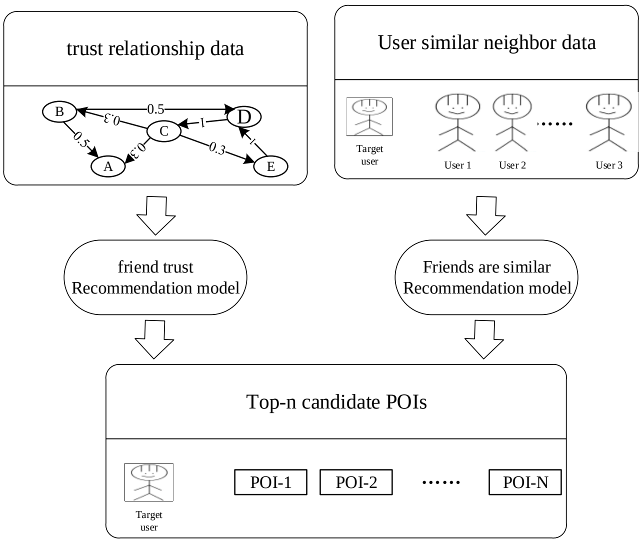


Figure 3. A framework for recommendation of POI based on social relationships

The candidate interest points of the two parts are fused, the formula is as follows:

$$ST_{i,j} = \alpha S_{i,j} + (1-\alpha)T_{i,j} \tag{7}$$

Where $S_{i,j}$ and $T_{i,j}$represent the probability of user check-in based on user similarity and trust relationship, respectively. The coefficient α ($0 \leq \alpha \leq 1$) is the adjustment factor. α =1 means that the recommendation system only considers the influence of the similarity between users and POI recommendation, and α =0 means that only the influence of peer trust on POI evidence is considered. In other cases, the recommender system considers the influence of two factors, user similarity and friend trust relationship.Its flow chart is as follows:

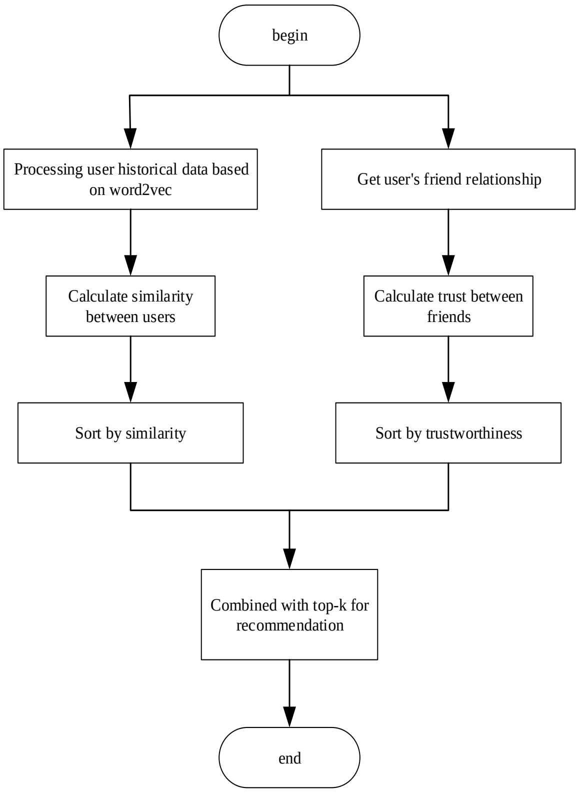


Figure 4. Overall flow chart

The model comprehensively considers social relationship information, effectively mines and combines friend similarity and friend trust factors, can recommend services for users more personalized and more accurately, and mine potential preferences that users have not visited but may be interested in.

4. Experiments

4.1 Experimental dataset

Table 1. Dataset

	Los Angeles	Washington	NYC
Amount of users	1278	932	1387
Number of POIs	1547	801	1475
Number of check-in records	25259	13720	28188
Number of friends	4466	2801	5625

The dataset used in this paper is the check-in dataset of three cities, Los Angeles (Los Angeles), Washington (Washington) and NYC (New York) in Gowalla (https://go.gowalla.com/). Includes check-in data from February 2009 to October 2010. The statistical details of the dataset are shown in Table 1.

In this paper, in chronological order, the first 80% of the data set is used as the training set, and the last 20% is used as the test set.

4.2 Evaluation indicators

Precision: The precision refers to the proportion of the actual number of users going to the recommendation results in the future to the total number of recommendations. It is assumed that after recommending K POIs to the target user, the target user goes to N POIs in the recommended POIs. The formula for calculating the accuracy rate is:

$$\text{precision@k} = \frac{\mathrm{N}}{\mathrm{K}} \tag{8}$$

Recall: The recall rate refers to the proportion of the actual number of users who will visit in the future in the recommendation results to the total number of POI interest that users visit in the future, reflecting the comprehensiveness of the recommendation. Assuming that the target user has visited M points of interest in the future, the point of interest may or may not be a recommended point of interest. The formula for calculating recall is:

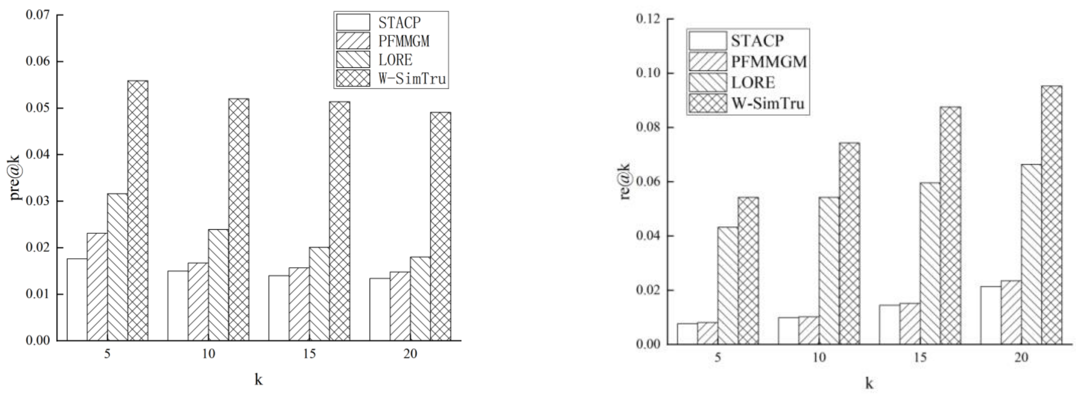
$$\text{recall@k=N/M} \tag{9}$$

The experiment tests the precision and recall of the recommendation results in the cases of K=5, 10 and 20, and the larger the value, the better.

4.3 Model comparison

In order to verify the accuracy of the model proposed in this paper, three models, STACP, PFMMGM and LORE are selected for performance comparison. The interest points recommended by each model take the precision and recall rates of top-k (k=5, 10, 15, 20). Compare.

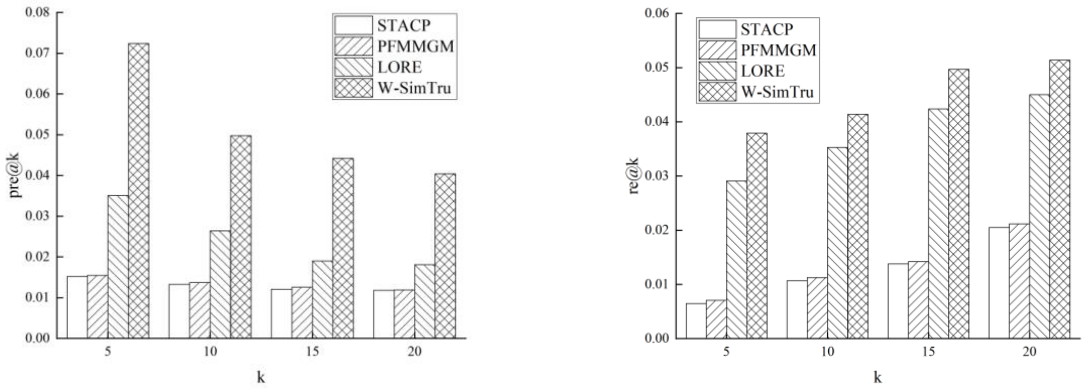
STACP [8]: A spatiotemporal activity center algorithm is proposed to more accurately model user behavior. The model is incorporated into a matrix factorization model in two different settings, static and dynamic, to demonstrate its effectiveness.



(a) Los Angeles-precision@k

(b) Los Angeles-recall@k

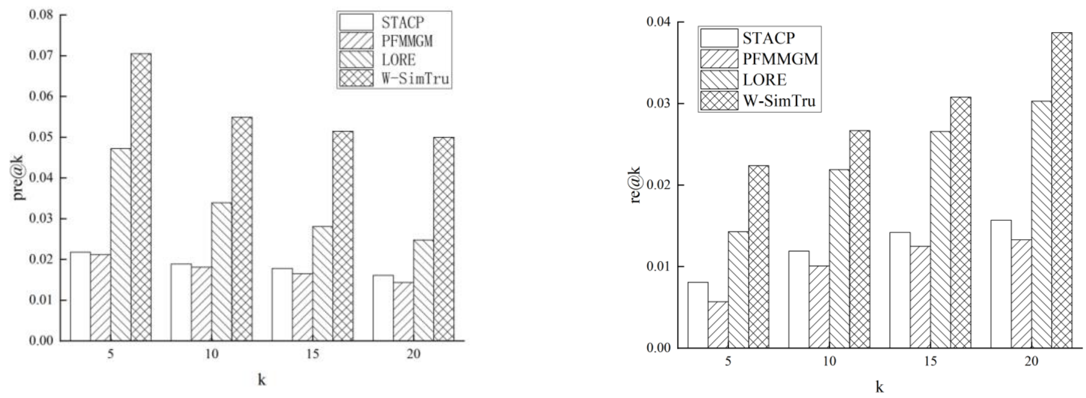
Figure 4. Los Angeles dataset



(a) Washington-precision@k

(b) Washington-recall@k

Figure 5. Washington dataset



(a) NYC-precision@k

(b) NYC-recall@k

Figure 6. NYC dataset

PFMMGM [9]: First obtain geographic influence by modeling user check-in frequency as a multi-center Gaussian model, and then incorporate social information and geographic influence into a matrix factorization framework for recommendation.

LORE[10]: predicts the probability of a user visiting a certain location through the additive Markov chain, and integrates the sequential influence with the geographical influence and social influence into a unified recommendation framework for recommendation.

W-SimTru: This paper proposes a recommendation model that combines user similarity and friend trust, and processes the dataset through word2vec to improve the performance of recommendation.

In the experiment, through multiple tunings, the optimal parameters of each model are selected, and then compared with the W-SimTru proposed in this paper. Verify the effectiveness of the model proposed in this paper.

Comparative experiments were performed on the Los Angeles, Washington, and NYC datasets, and the results are shown in Figure 4, Figure 5, and Figure 6.

The experimental results show that although the comparison model can recommend POI for users to a certain extent, the W-SimTru model proposed in this paper improves the recommendation by considering the user's own hobbies and the influence of friends in real life on their own decisions. performance, and is higher than the comparison algorithms in both precision and recall.

5. Conclusion

This paper proposes a point-of-interest recommendation model based on word2vec social relations. The social relationship includes not only the similarity relationship between two users, but also the trust relationship between friends. The model first processes the user's historical check-in data through word2vec, embeds the high-dimensional sparse data into low-dimensional dense check-in data, and then calculates the similarity of POI between two users through cosine similarity. At the same time, the trust degree between users' friends is calculated by TF-IDF, and the top-k points of interest are ranked by the combination of similarity and trust degree, and are recommended to users. Finally, experiments are carried out on real data sets, and compared with the three algorithms of STACP, PFMMGM and LORE, thus verifying the effectiveness of the W-SimTru algorithm proposed in this paper. In the future, we will try to combine this article with geographic location, hoping to further improve the performance of recommendation.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Grant: 61503169, 61802161) and Natural Science Foundation of Liaoning province (Grant: 2020-MS-291).

References

[1] Li, M., Zheng, W., Xiao, Y., Zhu, K., & Huang, W. (2021). Exploring temporal and spatial features for next POI recommendation in LBSNs. IEEE Access, 9, 35997-36007.

[2] Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS), 22(1), 5-53.

[3] Guo, Y., & Yan, Z. (2020). Recommended system: attentive neural collaborative filtering. IEEE Access, 8, 125953-125960.

[4] Liu, X., Liu, Y., Aberer, K., & Miao, C. (2013, October). Personalized point-of-interest recommendation by mining users' preference transition. In Proceedings of the 22nd ACM international conference on Information & Knowledge Management (pp. 733-738).

[5] Zhao, S., King, I., & Lyu, M. R. (2016). A survey of point-of-interest recommendation in location-based social networks. arXiv preprint arXiv:1607.00647.

[6] Xie, X. (2010, December). Potential friend recommendation in online social network. In 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing (pp. 831-835). IEEE.

[7] Ye, M., Yin, P., & Lee, W. C. (2010, November). Location recommendation for location-based social networks. In Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems (pp. 458-461).

[8] Wang, K., Wang, X., & Lu, X. (2021). POI recommendation method using LSTM-attention in LBSN considering privacy protection. Complex & Intelligent Systems, 1-12.

[9] Zhang, Z., Liu, Y., Zhang, Z., & Shen, B. (2019). Fused matrix factorization with multi-tag, social and geographical influences for POI recommendation. World Wide Web, 22(3), 1135-1150.

[10] Yang, D., Qu, B., Yang, J., & Cudre-Mauroux, P. (2019, May). Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In The world wide web conference (pp. 2147-2157).

[11] Yoon Y C, Lee J W. Movie recommendation using metadata based word2vec algorithm[C]//2018 International Conference on Platform Technology and Service (PlatCon). IEEE, 2018: 1-6.

[12] Church, K. W. (2017). Word2Vec. Natural Language Engineering, 23(1), 155-162.

[13] Rahmani, H. A., Aliannejadi, M., Baratchi, M., & Crestani, F. (2020, April). Joint geographical and temporal modeling based on matrix factorization for point-of-interest recommendation. In European Conference on Information Retrieval (pp. 205-219). Springer, Cham.

[14] Cheng, C., Yang, H., King, I., & Lyu, M. (2012). Fused matrix factorization with geographical and social influence in location-based social networks. In Proceedings of the AAAI conference on artificial intelligence (Vol. 26, No. 1, pp. 17-23).

[15] Zhang, J. D., Chow, C. Y., & Li, Y. (2014, November). Lore: Exploiting sequential influence for location recommendations. In Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 103-112). 103-112.

Author's biography

Li Yang-yang is currently enrolled in a Postgraduate degree program in Liaoning University of Technology, Liaoning Province, China. Major in Electronics and Communication Engineering, mainly studies the theory and application of signal and information processing. He is currently working on various algorithms for recommender systems.

Wang Ya-jun professor at Liaoning University of Technology, Jinzhou Liaoning Province, whose main research interests are theory and application of signal and information processing

Zhang Mi-yuan is currently enrolled in a Postgraduate degree program in Liaoning University of Technology, Liaoning Province, China. Major in information and communication engineering, mainly researching mobile communication and artificial intelligence. He is currently working on mmWave Massive MIMO systems.