# Learning Max-Margin GeoSocial Multimedia Network Representations for Point-of-Interest Suggestion

Zhou Zhao[1] Qifan Yang[1] Hanqing Lu[1] Min Yang[2] Jun Xiao[1*]
Fei Wu[1] Yueting Zhuang[1]
[1]College of Computer Science, Zhejiang University, China
[2]Tencent AI Lab, Shenzhen, China
{zhaozhou,bazinga,lhq110,junx,wufei,yzhuang}@zju.edu.cn,min.yang1129@gmail.com

## ABSTRACT

With the rapid development of mobile devices, point-of-interest (POI) suggestion has become a popular online web service, which provides attractive and interesting locations to users. In order to provide interesting POIs, many existing POI recommendation works learn the latent representations of users and POIs from users' past visiting POIs, which suffers from the sparsity problem of POI data. In this paper, we consider the problem of POI suggestion from the viewpoint of learning geosocial multimedia network representations. We propose a novel max-margin metric geosocial multimedia network representation learning framework by exploiting users' check-in behavior and their social relations. We then develop a random-walk based learning method with max-margin metric network embedding. We evaluate the performance of our method on a large-scale geosocial multimedia network dataset and show that our method achieves the best performance than other state-of-the-art solutions.

## CCS CONCEPTS

• **Information systems → Information retrieval**; *Recommender systems*;

## KEYWORDS

POI suggestion;network representation

## 1 INTRODUCTION

With the increasing popularity of mobile devices and Web 2.0 technology [10–12], geosocial multimedia network (GMN) has become a geographical web service that enables users to share their check-in behavior and photographs with geotagging [9, 15]. We have witnessed the popular GMN sites such as Foursquare and Gowalla. POI suggestion is an important component in GMN sites, which provides attractive and interesting locations to GMN users [3–5].

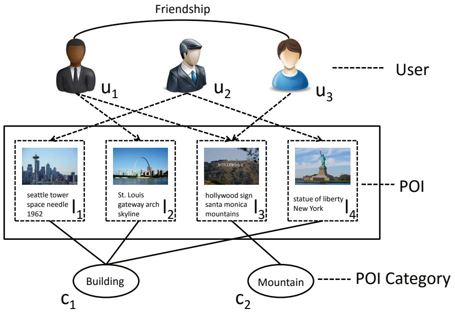


**Figure 1: GeoSocial Multimedia Network.**

The problem of POI suggestion has attracted considerable attention recently [2, 8]. Most of the existing works consider the POI suggestion problem as content-based location recommendation task, which learns the latent representations of users and POIs in GMN from users' check-in behaviors, and then suggest the right POIs to users [17]. Although existing POI suggestion methods have achieved promising performance, most of them still suffer from the insufficiency representation of POI contents and the sparsity of users' check-in data. Currently, most of existing POI suggestion methods [2, 8] learn the latent representations of users and POIs by approximating the binary user-POI rating matrix (each 0/1 in the matrix indicates whether a user has checked in at a POI) such that the max-margin metric is embedded in the representations for POI suggestion. However, the representation learning methods are mainly based on the hand-crafted features of POI contents such as POIs' tags and photographs. Recently, various multimodal deep representation methods [16] have been proposed for encoding the multimodal contents into the joint representation. Since the POIs' contents are multimodal, it is nature to employ the deep multimodal neural networks to learn their joint representations. On the other hand, we employ the categories of POIs for discriminative representation learning in order to gain better POI representation.

The sparsity of users' check-in data is also a challenging problem for POI suggestion. Each GMN user only visits a few POIs and thus the user-POI matrix is very sparse. Fortunately, with the prevalent of online social networks today, it is not difficult to find the social relations between GMN users. Following the popular homophily hypothesis, it is natural to assume that users' relations shows a strong evidence for their common preference. Thus, leveraging both users' social relations and their POI check-in behaviors are essential for tackling the sparsity problem.

[*] Corresponding author.

*SIGIR '17, August 07-11, 2017, Shinjuku, Tokyo, Japan*
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5022-8/17/08…$15.00
https://doi.org/10.1145/3077136.3080656

In this paper, we consider the problem of POI suggestion from the viewpoint of learning max-margin metric GMN network representation. We first introduce the multimodal neural networks for the POI representations under their categories. We then propose a random-walk based learning method with the introduced multimodal neural networks to jointly learn the representation of GMN users and POIs, such that the max-margin metric is implicitly embedded for POI suggestion. The main contributions of this paper are summarized as follows:

- Unlike previous studies, we introduce the problem of POI suggestion from the viewpoint of learning max-margin metric geosocial multimedia network representation. We learn the representations of users and POIs from GMN network with multimodal neural networks, such that the max-margin metric is implicilty embedded in the network representation for POI suggestion.
- We employ a random-walk based learning method with multimodal neural networks to learn the max-margin network representations for POI suggestion in GMN network. Our learning method is scalable for large-scale GMN networks which is easily parallelized.
- We evaluate the performance of our method using the dataset collected from the well-known geosocial multimedia network Gowalla and Flickr, and then show that our method achieves the best performance than other state-of-the-art solutions.

## 2 NOTATIONS AND RELATED WORK

Before reviewing previous work, we first introduce basic notions and terminologies for POI suggestion and network representation, then present the problem of POI suggestion. Given a set of POIs $L = \{\mathbf{l}_1, \ldots, \mathbf{l}_n\}$, we denote the collections of their associated photographs by $I = \{\mathbf{i}_1, \ldots, \mathbf{i}_n\}$ and associated tags by $T = \{\mathbf{t}_1, \ldots, \mathbf{t}_n\}$. We then denote the category of POIs by $C = \{c_1, \ldots, c_n\}$ where $c_i$ is the category of the $i$-th POI. We take the last hidden layer of the convolutional neural networks from POI's associated photographs as the visual representations by $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and the average of tag embeddings from POI's associated tags as semantic representations by $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$. We then denote the shared representations of the multimodal POIs by $Z = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$, where $\mathbf{z}_i$ is the fused representation vector of visual representation $\mathbf{x}_i$ and semantic embedding $\mathbf{y}_i$. We next consider the set of the user model representations by $U = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$, where $\mathbf{u}_i$ is the embedding vector for the $i$-th user model representation. We denote the friendship relation between users by matrix $S \in R^{m \times m}$, where the entry $s_{ij} = 1$ if the $i$-th user and the $j$-th user are friends, otherwise $s_{ij} = 0$. We then consider the check-in relation between users and POIs by $W \in R^{m \times n}$, where the entry $w_{ij} = 1$ if the $i$-th user checks-in the $j$-th POI, otherwise $w_{ij} = 0$.

We denote the propose the geosocial multimedia network by $G = (H, E)$ where the set of nodes $H$ consists of the joint representations of POIs, and users, the set of edges are composed of both the friendship relations $S$ and check-in relations $W$. We now illustrate a simple example of geosocial multimedia network in Figure 1. The $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ are users, $\mathbf{l}_1$, $\mathbf{l}_2$, $\mathbf{l}_3$ and $\mathbf{l}_4$ are POIs, and $c_1, c_2$ are the categories of POIs. The content of POIs is associated with photographs and tags. We construct the connection between POIs and users by their check-in behaviors. The friendship relation between users $\mathbf{u}_1$ and $\mathbf{u}_3$ is also illustrated in Figure 1.

Using the notations above, we define the problem of POI suggestion from the viewpoint of learning max-margin network representation as follows. Given the set of POIs $L$, users $U$, and the geosocial multimedia network $G$, we aim to learn the representations of POIs and users, such that the max-margin metric is implicitly embedded in the representations for POI suggestion.

## 3 POI SUGGESTION VIA LEARNING MAX-MARGIN MULTIMEDIA NETWORK REPRESENTATION

In this section, we first present the problem of POI recommendation from the viewpoint of max-margin heterogeneous network representation, and then introduce the max-margin network representation learning framework, illustrated in Figure 2.

We first construct the geosocial multimedia network by integrating the POIs' contents, users' check-in behaviors and their social relations, illustrated in Figure 2(a). We next sample the paths from GMN network as the context window for learning the vertex representation using the random-walk method in [13], shown in Figure 2(b). We introduce the proper multimodal neural network to learn the POI representations. We first employ the convolutional neural network for the visual representations of POIs' photographs. As there may be multiple photographs for each POI, we merge their representations by adding an additional max-pooling layers, illustrated in Figure 2(c). For the semantic representations of POIs' tags, we utilize the pre-trained word embedding and then add an additional mean-pooling layers to merge the tag representations, shown in Figure 2(c). We then set up the multimodal fusion layers to learn the joint POI representations, which connects the POI's visual representation and semantic representation into the same shared fusion space, and then add them together to obtain the activation of the multimodal fusion layer, given by

$$\mathbf{z}_i = g(\mathbf{Q}^v \mathbf{x}_i + \mathbf{Q}^s \mathbf{y}_i),$$

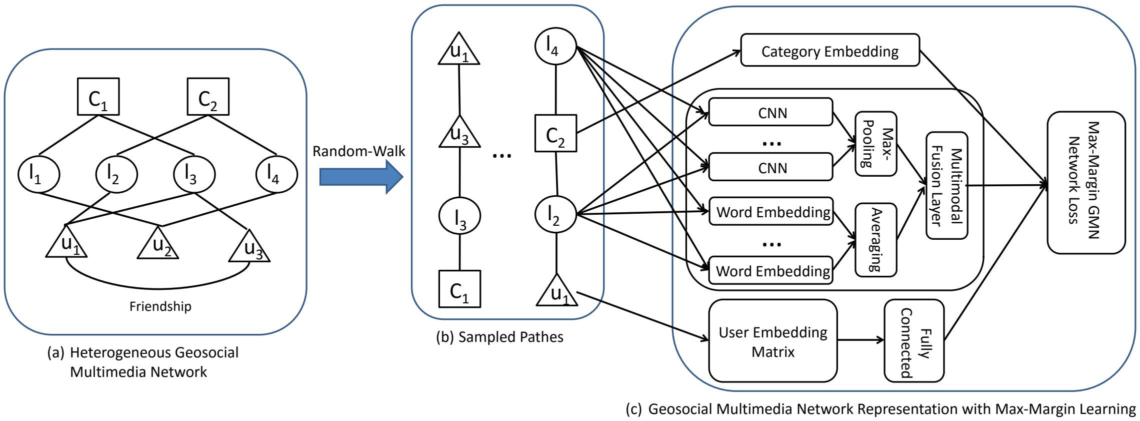
where the matrices $\mathbf{Q}^v$ and $\mathbf{Q}^s$ are the weights for projecting the visual representation and semantic representation into the shared space. We employ the element-wise scaled hyperbolic tangent function $g(\cdot)$ for merging the projected representations.

We now introduce the objective function for learning max-margin network representations based on the sampled paths and introduced multimodal neural networks. Given the sampled paths, illustrated in Figure 2(b), we consider them as the context window for vertex representation learning. For each vertex in the sampled path $h_i$, we now present the its loss function as follows:

$$l(h_i) = \begin{cases} \sum_{u_j \in P, W_{ji}=1} l_1(h_i, u_j) + \alpha \cdot l_2(h_i, c_i), & h_i \in L \\ \sum_{u \in W - h_i} \|u - h_i\|^2, & h_i \in U \end{cases}$$

where the max-margin metric loss $l_1(\cdot)$ is for users checking-in certain POIs and the max-margin metric loss $l_2(\cdot)$ is for learning the POI representation with its category information, and $\alpha$ is the tradeoff parameter. Thus, we instantiate the loss $l_1(\cdot)$ by

$$l_1(h_i, u_j) = \max(0, m_1 - \mathbf{h}_i^T \mathbf{u}_j), h_i \in L,$$



(a) Heterogeneous Geosocial Multimedia Network

(b) Sampled Pathes

(c) Geosocial Multimedia Network Representation with Max-Margin Learning

**Figure 2: The Framework of Heterogeneous Geosocial Multimedia Network Learning for POI Suggestion.**

and the loss $l_2(\cdot)$ by

$$l_1(h_i, c_i) = \max(0, m_2 - \mathbf{h}_i^T \mathbf{c}_i), h_i \in L,$$

where the hyperparameters $m_1$ and $m_2$ control the margin in the loss function.

We now present the training process of the proposed max-margin network learning method. We first start a random walker to sample the paths from the proposed GMN network, and then accumulate the training loss terms of the objective function. We next denote all the model parameters including the multimodal neural network parameters and the GMN network representations by $\Theta$. Therefore, the total loss function in our learning method is given by

$$\min_{\Theta} = \sum_{P} \sum_{h_i \in P} l(h_i) + \lambda \|\Theta\|^2, \tag{1}$$

where $\lambda$ is the tradeoff parameter.

## 4 EXPERIMENTS

In this section, we conduct several experiments to show the effectiveness of our method on POI suggestion. We collect the geosocial network from Gowalla [1] and obtain the multimedia POI contents from Flickr, and then construct the geosocial multimedia network. The dataset will be released for further study. Following the experimental setting of the POI suggestion problem in [2, 6–8], we employ the Precision@K and Recall@K to evaluate the effectiveness of the POI suggestion methods. We compare our proposed methods with other state-of-the-art methods CLR [7], STLR [6], CAPRF [2], GEOMF [8], CNNMSE [14] and DeepWalk [13] for the problem of POI suggestion.

Figures 3(a) and 3(b) illustrate the precision and recall of all algorithms using 80% training data, respectively. Tables 1, 2, 3 and 4 show the evaluation results of all methods using Precision@5, Precision@10, Recall@5 and Recall@10, respectively. The evaluation were conducted with different ratio from the training data from 20%, 40%, 60% to 80%. We report the average value of all methods using both Precision@K and Recall@K evaluation criteria. These experiments reveal a number of interesting points:

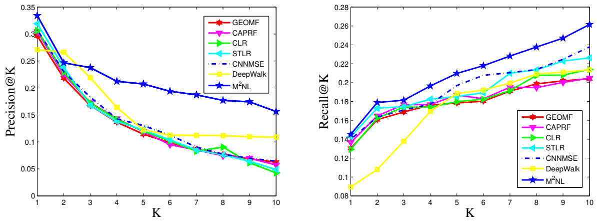


**Figure 3: Precision@K and Recall@K using 80% Data for Training.**

**Table 1: Experimental results on Precision@5**

| Method | Precision@5. | | | |
|---|---|---|---|---|
| | 80% | 60% | 40% | 20% |
| GEOMF | 0.1144 | 0.1020 | 0.0992 | 0.0874 |
| CAPRF | 0.1192 | 0.1038 | 0.0984 | 0.0815 |
| CLR | 0.1199 | 0.1049 | 0.0996 | 0.0905 |
| STLR | 0.1221 | 0.1000 | 0.0802 | 0.0734 |
| CNNMSE | 0.1304 | 0.0965 | 0.0859 | 0.0796 |
| DeepWak | 0.122 | 0.1147 | 0.1027 | 0.0831 |
| M$^2$NL | **0.2074** | **0.1731** | **0.1441** | **0.1291** |

**Table 2: Experimental results on Precision@10.**

| Method | Precision@10 | | | |
|---|---|---|---|---|
| | 80% | 60% | 40% | 20% |
| GEOMF | 0.0623 | 0.0506 | 0.0464 | 0.0413 |
| CAPRF | 0.0573 | 0.0523 | 0.0424 | 0.0409 |
| CLR | 0.0427 | 0.0414 | 0.0336 | 0.0309 |
| STLR | 0.0487 | 0.0446 | 0.0404 | 0.0387 |
| CNNMSE | 0.0644 | 0.0551 | 0.0501 | 0.0477 |
| DeepWak | 0.109 | 0.091 | 0.078 | 0.0675 |
| M$^2$NL | **0.1564** | **0.1375** | **0.1052** | **0.0787** |

- The CNNMSE method with deep neural networks based POI representation outperform other POI suggestion methods, which suggests that deep neural networks based POI

**Table 3: Experimental results on Recall@5.**

| Method | Recall@5 | | | |
|---|---|---|---|---|
| | 80% | 60% | 40% | 20% |
| GEOMF | 0.1784 | 0.1537 | 0.1364 | 0.1145 |
| CAPRF | 0.1869 | 0.1745 | 0.1609 | 0.1498 |
| CLR | 0.1798 | 0.1713 | 0.1577 | 0.1403 |
| STLR | 0.1853 | 0.1753 | 0.1516 | 0.1405 |
| CNNMSE | 0.1969 | 0.1674 | 0.1562 | 0.1461 |
| DeepWak | 0.1885 | 0.1532 | 0.1321 | 0.1035 |
| M$^2$NL | **0.2097** | **0.1726** | **0.1592** | **0.1468** |

**Table 4: Experimental results on Recall@10.**

| Method | Recall@10 | | | |
|---|---|---|---|---|
| | 80% | 60% | 40% | 20% |
| GEOMF | 0.2040 | 0.1947 | 0.1802 | 0.1723 |
| CAPRF | 0.2047 | 0.1983 | 0.1718 | 0.1585 |
| CLR | 0.2139 | 0.2051 | 0.1886 | 0.1677 |
| STLR | 0.2262 | 0.2091 | 0.1825 | 0.1697 |
| CNNMSE | 0.2378 | 0.2012 | 0.1913 | 0.1822 |
| DeepWak | 0.2135 | 0.216 | 0.1843 | 0.177 |
| M$^2$NL | **0.2615** | **0.2388** | **0.2221** | **0.1909** |

representation can enhance the performance of POI suggestion.

- The social POI suggestion method CAPRF also obtains good results, which demonstrates that the effect of social network is also critical for the problem.
- In all cases, our $M^2$NL method achieves the best performance, which shows that the max-margin GMN network representation method with deep POI representation and network learning can further improve the performance.

## 5 CONCLUSION

In this paper, we present the problem of POI suggestion from the viewpoint of learning max-margin GMN network representations. We propose the GMN network that exploits POIs' side information, users' check-in behaviors and their social relations for POI suggestion. We introduce the random-walk based learning method $M^2$NL with multimodal neural networks for learning GMN network representations, such that the max-margin metric is implicitly embedded in the representations for POI suggestion. The extensive experiments illustrate that our method can achieve better performance than several state-of-the-art solutions to the problem.

## ACKNOWLEDGMENTS

The work is supported by the National Natural Science Foundation of China under Grant No.61602405 and No.61572431. It is also supported by the Fundamental Research Funds for the Central Universities 2016QNA5015, Zhejiang Natural Science Foundation under Grant LZ17F020001, and the China Knowledge Centre for Engineering Sciences and Technology.

## REFERENCES

[1] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *SIGKDD*. ACM, 1082–1090.

[2] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. 2015. Content-Aware Point of Interest Recommendation on Location-Based Social Networks.. In *AAAI*. 1721–1727.

[3] Xiangnan He, Ming Gao, Min-Yen Kan, and Dingxian Wang. 2017. BiRank: Towards Ranking on Bipartite Graphs. *IEEE Trans. Knowl. Data Eng.* (2017), 57–71.

[4] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *WWW*. 173–182.

[5] Xiangnan He, Hanwang Zhang, Min-Yen Kan, and Tat-Seng Chua. 2016. Fast Matrix Factorization for Online Recommendation with Implicit Feedback. In *SIGIR*. 549–558.

[6] Bo Hu and Martin Ester. 2013. Spatial topic modeling in online social media for location recommendation. In *Proceedings of the 7th ACM conference on Recommender systems*. ACM, 25–32.

[7] Kenneth Wai-Ting Leung, Dik Lun Lee, and Wang-Chien Lee. 2011. CLR: a collaborative location recommendation framework based on co-clustering. In *SIGIR*. ACM, 305–314.

[8] Defu Lian, Cong Zhao, Xing Xie, Guangzhong Sun, Enhong Chen, and Yong Rui. 2014. Geomf: Joint geographical modeling and matrix factorization for point-of-interest recommendation. In *SIGKDD*. ACM, 831–840.

[9] Changzhi Luo, Bingbing Ni, Shuicheng Yan, and Meng Wang. 2016. Image Classification by Selective Regularized Subspace Learning. *IEEE Trans. Multimedia* (2016), 40–50.

[10] Liqiang Nie, Meng Wang, Yue Gao, Zheng-Jun Zha, and Tat-Seng Chua. 2013. Beyond Text QA: Multimedia Answer Generation by Harvesting Web Information. *IEEE Trans. Multimedia* (2013), 426–441.

[11] Liqiang Nie, Meng Wang, Zheng-Jun Zha, Guangda Li, and Tat-Seng Chua. 2011. Multimedia answering: enriching text QA with media information. In *SIGIR*. 695–704.

[12] Liqiang Nie, Shuicheng Yan, Meng Wang, Richang Hong, and Tat-Seng Chua. 2012. Harvesting visual concepts for image search with complex queries. In *ACMMM*. 59–68.

[13] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *SIGKDD*. ACM, 701–710.

[14] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In *NIPS*. 2643–2651.

[15] Meng Wang, Richang Hong, Guangda Li, Zheng-Jun Zha, Shuicheng Yan, and Tat-Seng Chua. 2012. Event Driven Web Video Summarization by Tag Localization and Key-Shot Identification. *IEEE Trans. Multimedia* (2012), 975–985.

[16] Meng Wang, Hao Li, Dacheng Tao, Ke Lu, and Xindong Wu. 2012. Multimodal Graph-Based Reranking for Web Image Search. *IEEE Trans. Image Processing* (2012), 4649–4661.

[17] Meng Wang, Xueliang Liu, and Xindong Wu. 2015. Visual Classification by -Hypergraph Modeling. *IEEE Trans. Knowl. Data Eng.* (2015), 2564–2574.