Recommend Places by Spatial and Non-Spatial Features

Mina Karimi*, Mohammad Saadi Mesgari*

* GIS Department, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran. mina-karimi@email.kntu.ac.ir, mesgari@kntu.ac.ir

Abstract. Places are described by users that make a huge number of textual contents. This work proposes a novel method for place recommender system based on the integration of semantic space and geographic space. To achieve this goal, salient features are modeled as directions in a domain specific semantic space. Finally, these directions and geographic distances can be used to rank places according to corresponding features.

Keywords. Semantic Space, Geographic Space, Place Recommendation.

1. Introduction

Concept of place, which is based on the experienced and observed space (rather than the commonly used geometric conception), is currently gaining attention (Purves et al. 2019; Wagner et al. 2020). The users of social networks provide a huge amount of textual content about places (Ballatore & De Sabbata 2020). Location-based recommender systems provide relevant suggestions to users by integrating location information (e.g. mobile GPS data), into algorithms (Quercia et al. 2010). These could include recommendations for hotels, restaurants, parks or other places or events near the user's location (Ye et al. 2011). Place recommendation would not recommend only near places but can be different according to the context. For instance, a recommender system of tourism should focus on attractiveness of places rather than their locations. In other words, these systems focus only on the general area of places, not on their exact locations. In addition, a place may be very attractive, but it is so far away that tourists will avoid

Published in "Proceedings of the 18th International Conference on Location Based Services (LBS 2023)", edited by Haosheng Huang, Nico Van de Weghe and Georg Gartner, LBS 2023, 20-22 November 2023 Ghent, Belgium.

This contribution underwent single-blind peer review based on the paper. https://doi.org/10.34726/5751 |

visiting it. Hence, this paper aims to propose a novel method for place recommender systems based on spatial and non-spatial features derived from textual contents. Our fully unsupervised approach contains three main steps requiring only a bag-of-words representation of online reviews of places as input.

In this paper we focus on information about what happens at a place or place functionality, in particular the activities that are associated with or afforded by a place. To achieve this goal, salient features would be modelled as directions in a domain-specific semantic space. Such domain-specific semantic spaces could be used to suggest items in recommender systems (Karimi et al. 2022, Abbasi & Alesheikh 2023). These directions can be used to rank objects (in this case, places) according to corresponding features. In addition, geographical coordinates will construct the geographic space to consider the distances between user and places.

2. Methodology

The general workflow of the proposed method is shown in Figure 1.

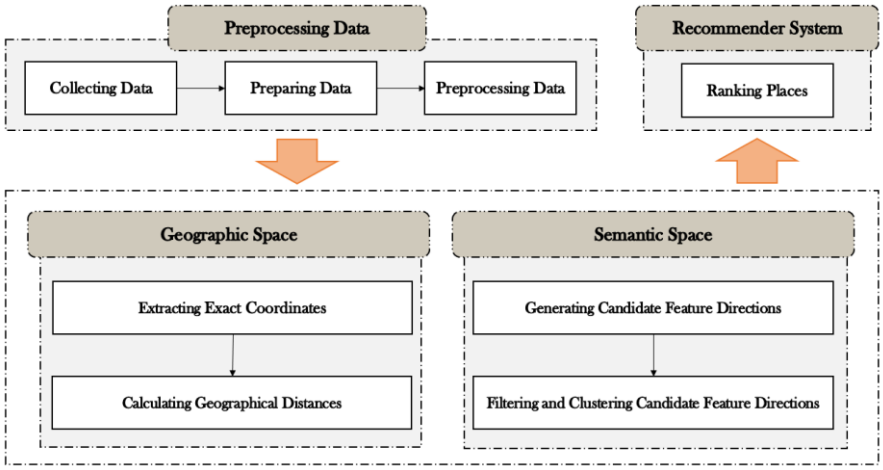

Fig. 1: The general workflow of the proposed method

2.1. Data

We are going to propose a method to online reviews of different places to recommend places based on their attractiveness and distance to user or another attraction. In the first step, places and their corresponding user reviews are extracted for New York City (NYC) by web crawling in Tripadvisor website, which were available in October 2020. Then, some preparing analysis is done to clean data and obtain useful dataset. Therefore, places without geographic locations, places out of the study area, duplicate places

and reviews, or places whose type is not "*attraction*" are removed. Afterward, to preprocess the user reviews, which are incredibly unstructured, each review is converted to lower case and tokenized. Then, punctuations and stop words are omitted. Finally, all tokens are lemmatized, and a bag of words (BoW) is created.

2.2. Constructing the Geographic and Semantic Spaces

In the second step, a cluster of words is used to define each feature (similar to LDA topic modeling but more flexible than it). Since there is not a priori knowledge about the most salient features, we consider all nouns, verbs and adjectives that are sufficiently frequent in the BoW representations of the places as candidate feature labels. These BoW representations of places are applied to learn a semantic space. This semantic space model salient features (e.g., how attractive a place is?) as directions. We will quantize the documents with respect to words by assigning the TF-IDF weighting scheme. Then, the vectors would be introduced into Multi-Dimensional Scaling (MDS) technique to construct a domain-specific semantic space. Each document in this semantic space is a point derived from MDS, where more similar points are located closer. Afterward, the most common words for topic Attraction are extracted by utilizing LDA to avoid sparse representation of the documents. Then, we will classify the space using a Logistic Regression binary classification to put a hyperplane in the space to distinguish those points (i.e., place documents) containing the given word and/or those lacking the word. The direction perpendicular to this hyperplane is the direction towards the given attribute (e.g., attractiveness). We will subsequently determine the most salient features in the considered domain, and their corresponding directions using the method from (Derrac & Schockaert 2015). Then, we evaluate the quality of the candidate feature directions using the accuracy of logistic regression classifier. For places with the accuracy of more than 50%, we rank places according to their distances to the hyperplane. The more the distance in the positive direction, the more the place is a tourist attraction. Hence, different places are ranked based on the probabilities.

Spatial information is significant in place recommendation systems among other non-spatial attributes. In the next step, geographical locations would be introduced to the model to consider the distances between user and places and the pairwise distances between places. Finally, these geographic and semantic spaces would be integrated to rank and recommend places according to the corresponding features (e.g., how attractive and close a place is).

3. Results and Discussions

After preparing data, around 482 attractions and 95661 user reviews are obtained for further analysis. We applied different python libraries such as Beautiful Soup, RE, NLTK, Gensim, Scikit-learn, and Matplotlib. Different MDS Dimensions are applied to find the best semantic space (5, 10, 15, 20, 50). The results demonstrate that D =15 will lead to better results, while increasing the dimension will not make significant improvement, but will require high computational cost. Table 1 shows the top 10 attractive places which are ranked based on the proposed method using the semantic space.

id	categoryType	Distance
1	[Sights & Landmarks]	81.44360911
2	[Muesums]	76.17219418
3	[Sights & Landmarks]	52.18535502
4	[Sights & Landmarks]	51.357416
5	[Other]	46.25181251
6	[Concerts & Shows]	45.07901105
7	[Other, Nature & Parks, Sights & Landmarks]	41.15674603
8	[Concerts & Shows]	40.97252379
9	[Fun & Games, Nature & Parks, Sports Camps & Clinics]	36.1009389
10	[Nature & Parks, Sights & Landmarks]	36.04665008

Table 1. The top 10 attractive places ranked in the semantic space.

It is worth mentioning that the authors are working on the combination of geographic space to semantic space to rank places based on the spatial and non-spatial features derived from travel websites. The thematic view of ranking places in semantic space is represented in Figure 1.

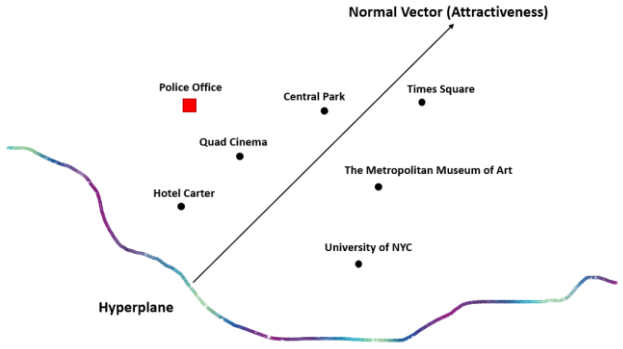

Figure 1. The thematic view of ranking places in the semantic space.

4. Conclusion

We introduced a method to rank places based on salient features. These features are modeled as directions in a domain-specific semantic space. Also, we are trying to integrate the geographic space to semantic space to design and implement a recommender system which is more applicable in tourism management. The proposed approach can help users to find places which are more attractive and closer to visit.

References

Abbasi, O. R., & Alesheikh, A. A. (2023). A Place Recommendation Approach Using Word Embeddings in Conceptual Spaces. *IEEE Access*.

Ballatore, A., & De Sabbata, S. (2020). Los Angeles as a digital place: The geographies of user-generated content. Transactions in GIS, 24(4), 880-902.

Derrac, J., & Schockaert, S. (2015). Inducing semantic relations from conceptual spaces: a data-driven approach to plausible reasoning. *Artificial Intelligence, 228*, 66-94.

Karimi, M., Mesgari, M. S., & Purves, R. S. (2022). A comparative assessment of machine learning methods in extracting place functionality from textual content. *Transactions in GIS*.

Karimi, M., Mesgari, M. S., Purves, R. S., & Abbasi, O. R. (2022). Modeling Salient Features as Directions for Place Recommender Systems. *Abstracts of the ICA*, *5*, 1-2.

Purves, R. S., Winter, S., & Kuhn, W. (2019). Places in information science. *Journal of the Association for Information Science and Technology, 70*(11), 1173-1182.

Quercia, D., Lathia, N., Calabrese, F., Di Lorenzo, G., & Crowcroft, J. (2010). *Recommending social events from mobile phone location data*. Paper presented at the 2010 IEEE international conference on data mining.

Stock, K., Jones, C. B., Russell, S., Radke, M., Das, P., & Aflaki, N. (2022). Detecting geospatial location descriptions in natural language text. *International Journal of Geographical Information Science*, *36*(3), 547-584.

Wagner, D., Zipf, A., & Westerholt, R. (2020). *Place in the GIScience community–an indicative and preliminary systematic literature review*. Paper presented at the Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19).

Ye, M., Yin, P., Lee, W.-C., & Lee, D.-L. (2011). *Exploiting geographical influence for collaborative point-of-interest recommendation*. Paper presented at the Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval.