

Spatio-Temporal Site Recommendation

Blinded for Double-Blind review

Abstract—Recommendation systems have become extremely common in recent years, and are utilized in a variety of areas to predict the “rating” or “preference” that a user would give to a point of interest (PoI), such as a restaurant, a hotel, or a bar. Such systems typically produce a list of recommendations by considering previous ratings of the user, as well as ratings of other users. Not every person rates every point of interest they visit. In this work, we want to explore the use of spatio-temporal data to improve recommendation systems: We postulate that spatio-temporal user data may indicate the liking or disliking of a point of interest. Clearly, if a user frequently visits the same PoI, stays at the PoI for long times, and is willing to travel a long distance to visit a PoI, that might indicate that user likes that PoI. Thus, we propose to extract user-PoI relation features from spatio-temporal trajectory data only. Using these features, we use out-of-the-box data mining and machine learning solutions, to estimate the popularity of a PoI. Our experimental evaluation shows, that the features extracted from spatio-temporal data able to accurately predict the popularity of a PoI, using ground-truth data from Yelp as a baseline.

I. INTRODUCTION

Modern technology to capture geo-spatial information produces a huge flood of individual trajectory data, coupled with a new user mentality of utilizing this technology to voluntarily share information. By mining this data, and thus turning it into actionable information, the McKinsey Global Institute [2] projects a “\$600 billion potential annual consumer surplus from using personal location data globally”. Towards this goal of making location data actionable, we propose to mine potential user ratings for location sites (e.g., restaurants, shops) from spatial-temporal data. User rating is critical in many applications, especially recommendation system. Techniques such as widely used collaborative filtering use known user rating information to estimate the interests of a user. However, in practice, the user-site rating matrix is usually highly sparse, meaning that there are not enough user rating information to perform such tasks. This is known as the cold-start problem. There are many reasons, for example, many users do not want to spend time to rate locations they visited, or locations such as shops do not have an efficient way to ask users for feedback.

To solve this problem, we propose to obtain implicit user-rating rather than explicit user ratings. Our approach uses trajectory data to find users that have likely visited a site. Since this data does not explicitly tell us whether the user likes that site, we propose to mine features from the users trajectory, which we intuitively expect to implicitly describe

whether the user likes that site. The features that we propose to obtain from trajectory data include:

- **The frequency of visits of a user.** If a user visits a site only once in their life, chances are that the user did not like the site enough to return. If the site is frequented often by the user, he seems to like it.
- **The length of stay of a user.** If a user stays at a restaurant or a hotel for a long time, that indicates that he likes the site.
- **The distance to their home base.** If a user is willing to take a long journey to reach a site, thus bypassing other, similar sites, that indicates that the user is subject to a strong attraction from that site, indicating that the user likes that site.

There are several advantages of using location data for location rating: (i) A potential solution to the cold-start problem, (ii) no user effort to capture their recommendation such as filling in rating forms, and (iii) it is based on user’s behavior, thus more objective and prone to alteration by fake-user-ratings and spam/bot-user-ratings.

Our approach becomes viable, due to the abundance of large open-sources collections of voluntarily contributed trajectory data, including the following data sources:

- **Location-Based Social Networks (LBSNs)** allows user to “Check-in” into a physical site such as a hotel, a restaurant or a metro station. Fairly large LBSN datasets, made anonymous, are available publicly. For instance, the FourSquare dataset used in [4] contains more than 30 million checkins and is available publicly. Such data explicitly includes the sites that a user has visited.
- **Geocoded Social Media Data:** is obtainable from public streaming APIs including for Twitter, Instagram, and Flickr. These data sources provide low-frequency trajectory data. Yet, microblogs and images are often published in sites of interest to a user, thus giving implicit information about the users site preferences.
- As part of the effort to create **Open-Street-Map (OSM)** [3], [1] road network data, users have been uploading GPS traces of their routes to the OSM site. While these routes are typically used to digitize the road network, the majority of routes is uploaded by pedestrians, and can be used as an indicator of sites that the corresponding user frequents. These trajectories are publicly available through the OSM API.

To describe our approach of site-recommendation using trajectory data, the rest of this work is organized as follows. We survey the state-of-the-art on site-recommendation and location-based recommendation systems in Section II. Then, we formally define the problem location-based site recommendation in Section III. Our solution, using deep-learning to bridge the gap from user-behaviour to user-site-recommendations is given in Section IV. Our solution is evaluated in Section VI, showing that our rating prediction for a restaurant is able to closely predict authoritative ground-truth site-ratings obtained from Yelp. We conclude our work in Section VII.

II. RELATED WORK

III. PROBLEM DEFINITION

IV. SPATIO-TEMPORAL USER-SITE FEATURE EXTRACTION

V. USER-SITE RATING PREDICTION

VI. EXPERIMENTAL EVALUATION

VII. CONCLUSIONS

REFERENCES

- [1] M. Haklay and P. Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, October-December:12–18, 2008.
- [2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. Big data: The next frontier for innovation, competition, and productivity. 2011.
- [3] Open street map. <http://www.openstreetmap.org>.
- [4] D. Yang, D. Zhang, L. Chen, and B. Qu. Nantotelescope: Monitoring and visualizing large-scale collective behavior in lbsns. *Journal of Network and Computer Applications*, 55:170–180, 2015.