UrbanHubble: Location Prediction and Geo-Social Analytics in LBSN

Roland Assam(✉), Simon Feiden, and Thomas Seidl

RWTH Aachen University, Aachen, Germany
{assam,seidl}@cs.rwth-aachen.de, simon.feiden@rwth-aachen.de

Abstract. Massive amounts of geo-social data is generated daily. In this paper, we propose UrbanHubble, a location-based predictive analytics tool that entails a broad range of state-of-the-art location prediction and recommendation algorithms. Besides, UrbanHubble consists of a visualization component that depicts the real-time complex interactions of users on a map, the evolution of friendships over time, and how friendship triggers mobility.

1 Introduction

The volume of data generated from human social interactions in Location-Based Social Networks (LBSN) is breathtaking. Such data encapsulates all visited locations and mimics the identity, behaviors, and affiliations of an individual or group. This has fueled enormous research interests to study location-based social interactions or group dynamics. One profound user behavior that has emerged during mobile social networking is the generation of *check-in*. Check-in is a phenomena whereby a person deliberately broadcasts her current location to a group of friends in an LBSN.

Numerous location prediction techniques have been proposed. To the best of our knowledge, there is no platform that consists of a broad array of innovative state-of-the-art location prediction techniques such as [1,3,5,6,8,9]. The availability of such a framework would assist researchers to quickly compare and evaluate state-of-the-art prediction techniques. Thus, saving their time and allowing them to focus more on the new techniques they aspire to develop.

Towards this end, we were motivated to create UrbanHubble[1], an innovative LBSN predictive analytics tool, which entails a broad spectrum of state-of-the-art LBSN prediction algorithms. Specifically, the algorithms include [1,3,5,6,8,9]. While Spot [4] also provides a platform to analyze LBSN, it consists of three algorithms. In contrast, we provide more algorithms than [4] and most importantly, our framework contains the most recent or relevant location prediction techniques. In addition to the aforementioned algorithms, UrbanHubble consists of a visualization component that shows the real-time complex social interactions of users on a map, the evolution of friendships over time, and how friendship triggers mobility or vice-versa.

[1] http://dme.rwth-aachen.de/en/urbanhubble

A. Bifet et al. (Eds.): ECML PKDD 2015, Part III, LNAI 9286, pp. 329–332, 2015.
DOI: 10.1007/978-3-319-23461-8_41

While UrbanHubble is primarily intended for researchers, its visualization interface can also be used across an array of industries such as in location-based advertisement, where customers behaviors' strongly depend on the location context, and advertisers are interested to efficiently identify patterns to hyper target such customers, or in urban planning or for traffic monitoring. Given the importance and enormous potentials of LBSN research, we believe these use-cases connote the demand for UrbanHubble.

2 UrbanHubble Tool

In this section, we provide a detailed description of the UrbanHubble predictive analytics Java platform. We provide four real datasets that are publicly available. They include, two versions of the LBSN Gowalla[2] check-in datasets (where one Gowalla dataset has category information while the other does not), the Brightkite LBSN dataset[3] and the non-LBSN San Francisco taxi cab[4] dataset.

2.1 Architecture

The architecture of UrbanHubble is depicted in Figure 1. It entails 4 layers, namely, the Persistence, Scalability, Predictive Engine and User Interface layer.

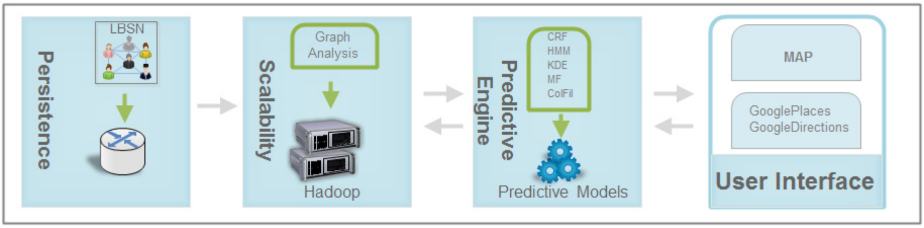


Fig. 1. Urban Hubble Architecture.

Persistence Layer: is home to the previously alluded datasets. To determine important social-relationship properties of users, we represent the collection of users as a simple graph G.

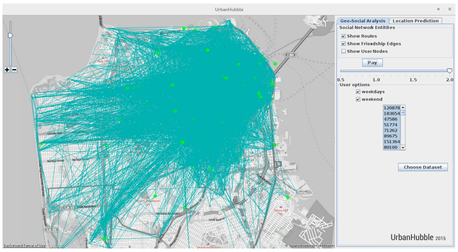
Scalability Layer: The number of check-ins from the Gowalla and Brightkite datasets exceeds a million. To rapidly analyze the check-ins correlations and the complex interconnections between the users in the huge graph G, a scalability layer is required. The scalability layer entails several Hadoop clusters to process G during Matrix Factorization and other exhaustive computations.

Predictive Engine Layer: consists of a wide variety of state-of-the-art location prediction algorithms that use Hidden Markov Model (HMM), Conditional

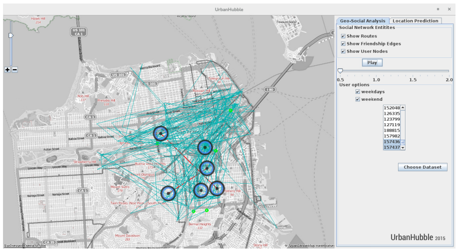
[2] https://snap.stanford.edu/data/loc-gowalla.html
[3] https://snap.stanford.edu/data/loc-brightkite.html
[4] http://cabspotting.org/

Random Fields (CRF), Matrix Factorization (MF), Collaborative Filtering and Kernel Density Estimation (KDE). Specifically, the UrbanHubble framework consists of six algorithms. They include [1,3,5,6,8,9]. Due to space constraints, we briefly describe a few of the algorithms packaged in UrbanHubble. [1] is our predictive model that uses Lipschitz exponent and CRF to determine the top-k future locations. [5] utilizes matrix factorization and KDE to predict and recommend the top-k locations.

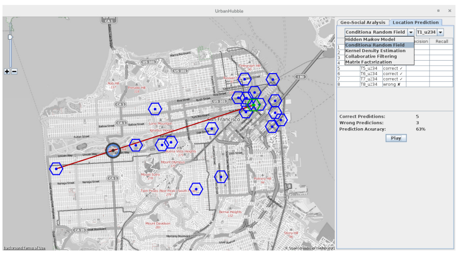
User Interface: The user interface consists of two tabs. They include, the Geo-Social Analysis tab and the Location Prediction tab as illustrated in Figure 2a. The former shows the real-time dynamics between mobility and friendship at a city resolution, while the later illustrates the detailed trajectory paths taken by users. The user interface of the *Geo-Social Analysis* tab provides a researcher the possibility to select an algorithm and the option to analyze geo-social mobility on weekdays or weekends. Furthermore, there is an option to chose either to analyze only trajectories (i.e., Show Route) or the social interactions between users. After selecting the desired configurations, the Play button can be clicked to run the selected algorithm. As the algorithm runs, the movements of users and their social connections are displayed on the map. On this tab, the small green balls in Figure 2a correspond to the end destinations of users, the dark cyan lines are the routes, while the red edges represent the friendships between users. In addition, the large circles denote nodes or users.



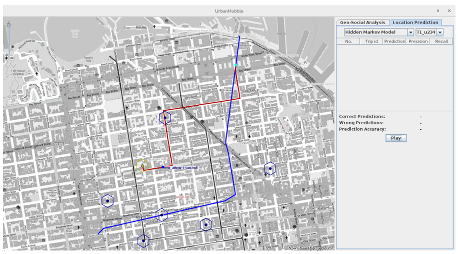
(a) LBSN in San Francisco.



(b) Geo-Social Analysis.



(c) Check-in Predictions



(d) Prediction with TaxiCab.

Fig. 2. Urban Hubble User Interface.

On the *Location Prediction* tab, each hexagon (e.g., in Figure 2c and Figure 2d) is a candidate next location. If the category information is absent from the dataset, to run the WhereNext [8] algorithm, UrbanHubble queries Google Places API to determine the category of each cell as shown in Figure 1. As mentioned earlier, we provide a default check-in dataset with category. After an algorithm runs to completion, the results of the algorithm are displayed as shown in Figure 2c using the precision, recall and accuracy evaluation measures. Figure 2d shows a scenario where the San Francisco TaxiCab dataset is used.

2.2 Related Works

[2] presented a reachability-based predictive model to predict check-in locations for distant-time queries. Our work differs from [2,4] since the algorithms packaged in our framework are different from theirs. Besides, our framework has more visualization functionalities that is not only limited to prediction analysis but also depicts geo-social interaction and group dynamics. MoveMine 2.0 [7] is a framework that focuses on trajectory clustering.

3 Conclusions

We propose an innovative LBSN predictive analytics java framework called UrbanHubble, which consists of a wide variety of state-of-the-art prediction algorithms. UrbanHubble has a visualization component and would be beneficial to researchers who intend to evaluate recent innovative LBSN prediction and recommendation techniques.

References

1. Assam, R., Seidl, T.: Check-in location prediction using wavelets and conditional random fields. In: ICDM 2014, pp. 713–718 (2014)
2. Chiang, M.-F., Lin, Y.-H., Peng, W.-C., Yu, P.S.: Inferring distant-time location in low-sampling-rate trajectories. In: KDD (2013)
3. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: KDD 2011, pp. 1082–1090 (2011)
4. Kong, L., Liu, Z., Huang, Y.: Spot: locating social media users based on social network context. In: VLDB 2014, pp. 1681–1684 (2014)
5. Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., Rui, Y.: Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In: KDD 2014, pp. 831–840 (2014)
6. Lichman, M., Smyth, P.: Modeling human location data with mixtures of kernel densities. In: KDD 2014, pp. 35–44 (2014)
7. Wu, F., Lei, T.K.H., Li, Z., Han, J.: Movemine 2.0: mining object relationships from movement data. In: VLDB 2014, pp. 1613–1616 (2014)
8. Ye, J., Zhu, Z., Cheng, H.: What's your next move: user activity prediction in location-based social networks. In: SDM 2014, pp. 171–179 (2013)
9. Ye, M., Yin, P., Lee, W.-C., Lee, D.-L.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: SIGIR (2011)