Tackling Cold-Start Users in Recommender Systems with Indoor Positioning Systems

Emanuel Lacic
Graz University of Technology
Graz, Austria
elacic@know-center.at

Dominik Kowald
Know-Center
Graz, Austria
dkowald@know-center.at

Matthias Traub
Know-Center
Graz, Austria
mtraub@know-center.at

Granit Luzhnica
Know-Center
Graz, Austria
gluzhnica@know-center.at

Joerg Simon
Know-Center
Graz, Austria
jsimon@know-center.at

Elisabeth Lex
Graz University of Technology
Graz, Austria
elisabeth.lex@tugraz.at

ABSTRACT

In this paper, we present work-in-progress on a recommender system based on Collaborative Filtering that exploits location information gathered by indoor positioning systems. This approach allows us to provide recommendations for "extreme" cold-start users with absolutely no item interaction data available, where methods based on Matrix Factorization would not work. We simulate and evaluate our proposed system using data from the location-based FourSquare system and show that we can provide substantially better recommender accuracy results than a simple MostPopular baseline that is typically used when no interaction data is available.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data mining*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*

Keywords

cold-start; IPS; beacon; collaborative filtering; FourSquare

1. INTRODUCTION

One of the main challenges in recommender systems is the cold-start problem which is defined by so-called cold-start users who have not a single or only very few item interaction data available (e.g., ratings). In order to tackle this problem, systems like MovieLens typically provide interaction surveys where a new user has to fulfill a predefined number of interactions before recommendations can be calculated. However, users are often annoyed by such surveys or find it hard to immediately come up with a representative list of item ratings to fill them out.

Another way to address cold-start users is to utilize algorithms based on Matrix Factorization. Although these methods are able to provide reasonable results when a minimum number of user-item interactions is available (e.g., three ratings, see [2]), they fail in "extreme" cold-start settings where there are no item interactions. In such cases, recommender systems typically make use of unpersonalized methods such as providing the overall most popular items in a system. Since recommendations should be personalized in order to support users in the most efficient way, we investigate the usefulness of an additional data source in order to tackle such "extreme" cold-start users with no item interactions at all. One opportunity in this respect would be to make use of the ever increasing trend of providing mobile applications to help users navigate through different kinds of public areas, such as shopping centers or scientific conferences. These applications can easily acquire a user's location information using indoor positioning systems (IPS) [1] to automatically collect location-based item interaction data with no need for any explicit user action (e.g., a click).

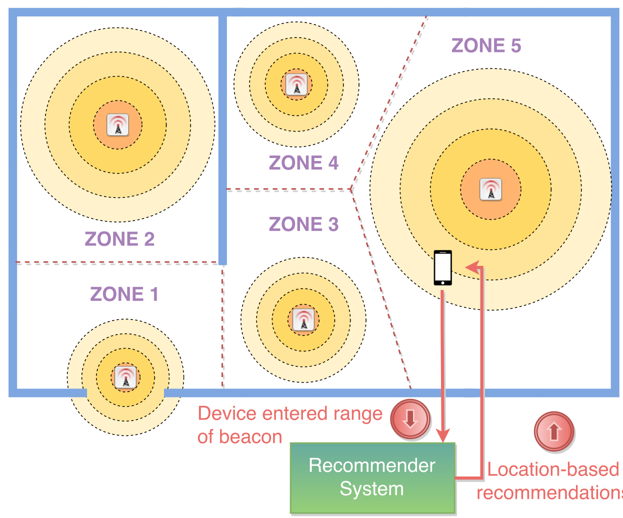

Figure 1: Example of a public area (e.g., shopping center or academic conference) with five zones where to each zone a beacon is attached. When a device (e.g., a smartphone) enters a zone, the data is stored in our recommender system and can then be utilized for location-based recommendations.

We make use of a user's location data gathered via IPS technology by proposing a novel recommender system, which utilizes the user-based Collaborative Filtering approach. Thus, we compute the similarity between two users based on (i) raw location data and (ii) by creating a user-location network that connects users who visited the same location during the same day and hour. The preliminary results of our evaluation based on FourSquare data show that our proposed approach provides substantially better recommender accuracy results than a simple Most Popular baseline that is typically used when no user-item interaction data is available.

RecSys 2015 Poster Proceedings, September 16–20, 2015, Austria, Vienna.

2. PROPOSED APPROACH

Tracking User Locations. There exists a number of easily attainable technologies, or indoor positioning systems (IPS), to track indoor locations. Among them, BLE (Bluetooth Low Energy) beacons have gained importance and popularity, especially after Apple introduced the iBeacon protocol[1]. Beacons are basically a small piece of hardware that can be easily attached to e.g., a wall and transmit a broadcast to every smartphone or a tablet within its reach. Beacons are especially applicable for recommendation tasks since they provide both indoor localization and proximity sensing at low cost and low energy. In our case, we have a public area such as a shopping center or an academic conference which is divided into several zones. A zone is an abstract location represented by a beacon with a given radius (see Figure 1), containing a certain set of co-located items (e.g., products or venues), preferably related to each other. The transmission power of the broadcast signal should be tuned to match the respective physical area of the corresponding zone. However, it should be considered that errors in approximating the distance increase with the size of the signal distance [5].

Recommender System. Our IPS-based recommender system relies on user-based Collaborative Filtering. We calculate the similarity between users u and v either by using the Jaccard's Coefficient: $sim(u, v) = \frac{|\Delta(u) \cap \Delta(v)|}{|\Delta(u) \cup \Delta(v)|}$ on their raw location data (denoted by $\Delta(u)$ and $\Delta(v)$, respectively), or by constructing a location network where ties between two users are existent if they visited the same location within the same day and hour. On the constructed location network in which $\Gamma(u)$ denotes the location-based neighbourhood of user u, we apply related similarity metrics: *Neighbourhood Overlap*: $sim(u, v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u)| + |\Gamma(v)|}$, and a refinement proposed as *Adamic Adar*, which adds weights to the links (since not all neighbours in a network have the same tie strength): $sim(u, v) = \sum_{z \in \Gamma(u) \cap \Gamma(v)} \frac{1}{log(|\Gamma(z)|)}$ (see [3] for the complete formalism).

From a technical perspective, we utilized the recommender framework presented in [4] to implement and evaluate our approach.

3. EVALUATION

Experimental Setup. We evaluated our IPS-based recommender approach with respect to nDCG (see e.g., [2]) using the FourSquare dataset provided by [6]. We chose this dataset since FourSquare best simulates our setting of a public area (e.g., shopping center or academic conference) that can be tracked with IPS technology. Our primary focus lies on users with no item interaction data in the training set, and our approach recommends up to 10 items (i.e., venues in the FourSquare setting). Thus, we extracted all users that interacted with 10 items (= 2,783 out of 2,153,471 users) and put these interactions into the test set to be predicted. This ensures that each of these users is an "extreme" cold-start user. In order to finally evaluate the effectiveness of our approach, we compared it to a standard MostPopular baseline, which is the most intuitive way to provide recommendations when no item interaction data is available.

Preliminary Results. The preliminary results of our evaluation are shown in Figure 2 in form of a nDCG plot. The results indicate that all three location-based CF approaches outperform the MostPopular baseline which is the standard method for handling users with no item interaction data available. Regarding the location-based algorithms, the two methods based on a user-location network, which connects users who visited the same location during a defined period of time, provide higher nDCG estimates than the method solely based on the raw location data. The overall best results are reached by the location network-based approach using the Adamic Adar metric with a nDCG@10 value of nearly 15%.

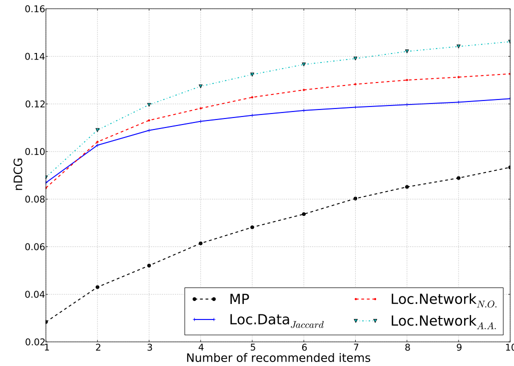

Figure 2: nDCG plot for "extreme" cold-start users in the FourSquare dataset showing that all three location-based CF algorithms outperform the MostPopular baseline.

4. CONCLUSION AND FUTURE WORK

In this paper, we have presented work-in-progress on a novel recommender system that tackles "extreme" cold-start users with indoor positioning systems (i.e., beacon technology). Furthermore, we have shown that our approach outperforms the MostPopular baseline in an experiment on FourSquare data. One limitation of our experiment is that it only simulates our approach but it clearly shows the potential of it. Thus, as a next step, we will conduct a large-scale user study to evaluate our approach in a real setting by including it into the i-KNOW Conference Assistant[2] during the next i-KNOW conference in October 2015. This system will not only recommend talks and events but also papers and people according to a user's interests and visited indoor locations.

Additionally, we plan to use the accelerometer and gyroscope sensor to detect the direction of a user in relation to the location of items and try to exploit this for recommendations. We aim to differentiate between cases where a user randomly (i.e., without a specific intention) passes through a zone versus cases where a user visits a zone and is looking at an item for a longer time or at closer distance. Hence, we can prevent spamming the user with recommendations while hassling through a public area.

Acknowledgments: The authors would like to thank Matthias Heise for helpful comments on this work. This work is supported by the Know-Center and the EU-IP Learning Layers (Grant Agreement: 318209).

5. REFERENCES

[1] J. D. Cai. Business intelligence by connecting real-time indoor location to sales records. In *WAIM '14*. Springer.

[2] D. Kluver and J. a. Konstan. Evaluating recommender behavior for new users. *Proc. of RecSys '14*.

[3] E. Lacic, D. Kowald, L. Eberhard, C. Trattner, D. Parra, and L. B. Marinho. Utilizing online social network and location-based data to recommend products and categories in online marketplaces. In *Mining, Modeling, and Recommending 'Things' in Social Media*. 2015.

[4] E. Lacic, D. Kowald, and C. Trattner. Socrecm: A scalable social recommender engine for online marketplaces. In *Proc. of HT '14*.

[5] P. Martin, B.-J. Ho, N. Grupen, S. Muñoz, and M. Srivastava. An ibeacon primer for indoor localization: Demo abstract. In *Proc. of BuildSys '14*.

[6] M. Sarwat, J. Levandoski, A. Eldawy, and M. Mokbel. Lars*: An efficient and scalable location-aware recommender system. *Knowledge and Data Engineering, IEEE Transactions on*.

[1]https://developer.apple.com/ibeacon/

[2]http://is.gd/EdMYCN