How Long to Stay Where? On the Amount of Item Consumption in Travel Recommendation

Linus W. Dietz
Department of Informatics
Technical University of Munich
Garching, Germany
linus.dietz@tum.de

Wolfgang Wörndl
Department of Informatics
Technical University of Munich
Garching, Germany
woerndl@in.tum.de

ABSTRACT

Recommender systems could benefit from not only recommending the most fitting items, but also in what quantity the user should consume them. For example, a personalized travel recommender system could indicate not just which city one should travel to, but also how much time to spend there. We present a data-driven solution to this problem based on mining trips from location-based social networks. To determine the recommended duration of stay at a destination, we consider how long travelers typically stay at different cities and how much time the current user generally spends visiting cities.

KEYWORDS

recommender systems; user modeling; travel recommendation

INTRODUCTION

Recommender systems research is mostly concerned with predicting the ratings for items of an active user, determining an optimal ranking of items, and presenting top-ranked items in an appealing way. This challenge of finding the "best" item according to any metric is essential in virtually all recommender system domains. However, items can also be recommended multiple times, such as if

ACM RecSys 2019 Late-breaking Results, 16th-20th September 2019, Copenhagen, Denmark

favorite artists appear repeatedly in a long music playlist. In this case, the assumption is that an item should be watched, listened to or experienced as a whole, not only parts of it.

In some scenarios, it is important to decide not just which items should be recommend, but how much of an item should be consumed. For example, a destination recommender should not only recommend where to go, but also the optimal duration of one's stay at each location. The duration can vary depending on the relevance of the item and other domain-related factors, such as the type of traveler [4]. Furthermore, items may be recommended multiple times within a travel package [8]. For example, a recommendation regarding the perfect day at an amusement park might call for riding on the same attraction multiple times.

In this paper, we examine the problem of determining the personalized amount of recommended item consumption in recommender systems, since previous approaches have not solved this problem convincingly. We then present a way to derive the duration of stays in the domain of destination recommendation. Finally, we discuss the generalizability of our method and draw conclusions.

RELATED WORK

There is very limited related work with regard to determining the amount of item consumption in recommender systems for travel and tourism.

Melià-Seguí et al. have investigated the typical duration of stays for tourists visiting different point-of-interest categories using a Foursquare data set; for example, they considered the average amount of time that users spent in restaurants [7]. Google Maps also presents information on how much time visitors spent at selected venues in its search results. However, this information represents only the duration of visits to individual locations or categories of locations; it cannot be used directly to construct recommendations on how long to stay in a city or travel region.

There are several approaches to recommending travel packages, such as the Tourist-Area-Season Topic (TAST) Model [6]. The idea underlying this model is to analyze features of travel packages with regard to their item and user representations, which can then be utilized in a recommender system. In this and similar work, features such as seasonality and item prices are often taken into account, but the duration of stay is either fixed and predetermined, or not considered at all. A related problem is to combine several destinations in a single composite trip. Since travelers' time availability and budget are usually constrained, this recommendation problem can be modeled as a knapsack problem with a scoring function that balances the benefit and cost of items within the package. Herzog and Wörndl have presented an approach to scoring travel regions based on user preferences and then combining them into a longer trip [5]. The score of a region is gradually decreased on a weekly basis, so different regions with lower initial scores may be added to the knapsack. However, this adjustment of the duration of stay is very coarse and not adapted to item or user characteristics in more detail.

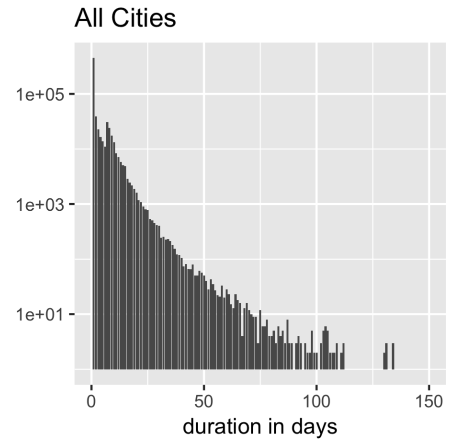

Figure 1: Distribution of the durations of blocks of trips in all 3,938 cities

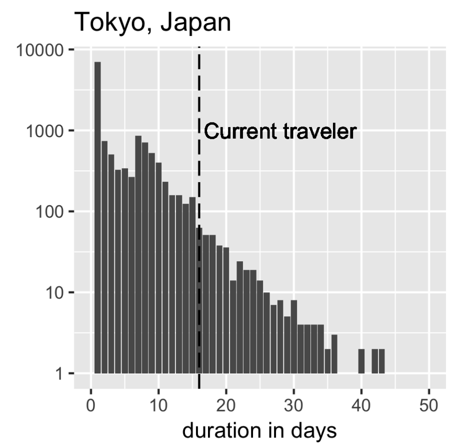

Figure 2: Distribution of the durations of blocks of trips in Tokyo, Japan

When recommending a sequence of travel-related items, such as an itinerary for a city visit, the problem of how much time to spend at individual locations arises as well. For example, De Choudhury et al. have analyzed Flickr photo streams to reconstruct paths of tourists in a city [1]. This information is useful for creating an interesting itinerary once a tourist has already selected a city to visit, but it does not tackle the problem of where to go on a trip or for how long. The determination of the duration of stay is an open research problem [2], which could be resolved through mobility analysis of traveler data [3]. To the best of our knowledge, no existing approach adequately addresses the problem raised here.

DERIVING THE DURATION OF STAY IN A DESTINATION RECOMMENDER SYSTEM

Having analyzed the related work, we will now sketch our ideas as to how to resolve the problem in the domain of destination recommendation. Our proposed solution addresses the question of making personalized recommendations regarding the duration of a tourist's visit to a city by considering two factors: the typical time that all tourists spend in that city and the particular user's average length of stay at a given destination. Initially, we need to know the distribution of the durations of people's stay at a destination, since there can be substantial differences between destinations as to how long one needs to explore it. For example, a smaller city can be covered within a day or two, whereas a major metropolis might require more time. The second aspect is the pace at which the particular traveler visits destinations. Some tourists want to immerse themselves deeply into a culture and therefore stay at each location for a longer time, whereas others want to visit as many different places as possible during their holidays. To quantify these behaviors, we need a database of previous trips to establish a distribution of how long people stay at a specific destination, such as a country or a city.

We employ our previously proposed approach to mine trips from a data set stemming from Foursquare [3], a location-based social network (LBSN), where people can check in at venues all over the world. However, the analysis presented in that paper is at a country-level granularity, whereas we look at the duration of stays at the city level. Using a Foursquare data set of 33,278,683 check-ins by 266,909 users [9], we mine 223,688 domestic and 10,963 international trips, requiring a minimum duration of seven days to mitigate the confounding effect of short business trips. These trips are further segmented into blocks, which are consecutive check-ins at the same municipality with over 15,000 inhabitants. The trips have a mean value of 2.944 blocks, resulting in a total of 690,897 blocks for further analysis. The bar plot in Figure 1 shows the distribution of the durations of all blocks, regardless of the city. The logarithmized counts show a bimodal distribution, with most blocks being one day long and another peak at seven days. This second peak can be attributed our decision to set the minimum duration of the whole trip at seven days.

The next step is to determine the pace at which our particular user typically travels, i.e., the distribution of the duration the individual traveler's past blocks. To obtain this information, we can either ask the traveler to provide some information about past trips directly, or we can request access to the individual's mobility patterns from her profile on a LBSN. Once we have this information about past trips, we can derive the user's pace by comparing it to the quantiles of all other travelers who have visited the same destinations. This essentially establishes a collaborative filtering method to derive the duration of stays from actual user behavior.

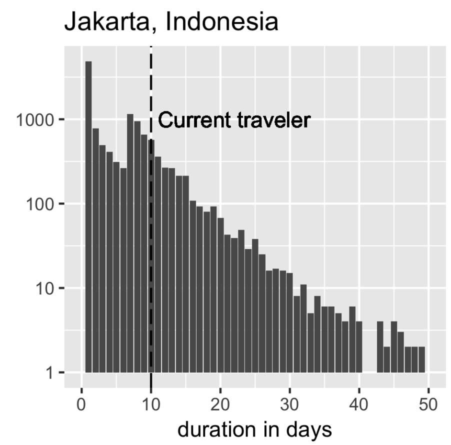

Figure 3: Distribution of the durations of stay of blocks in Jakarta, Indonesia

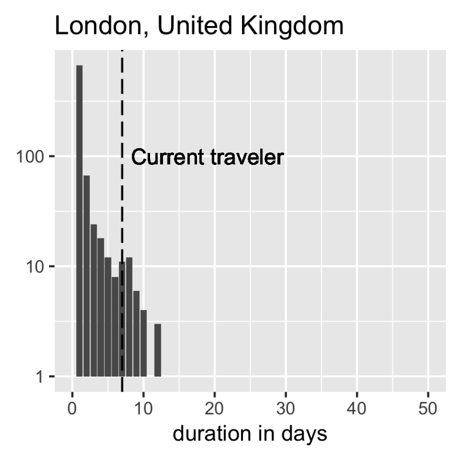

Figure 4: Distribution of the durations of stay of blocks in London, United Kingdom

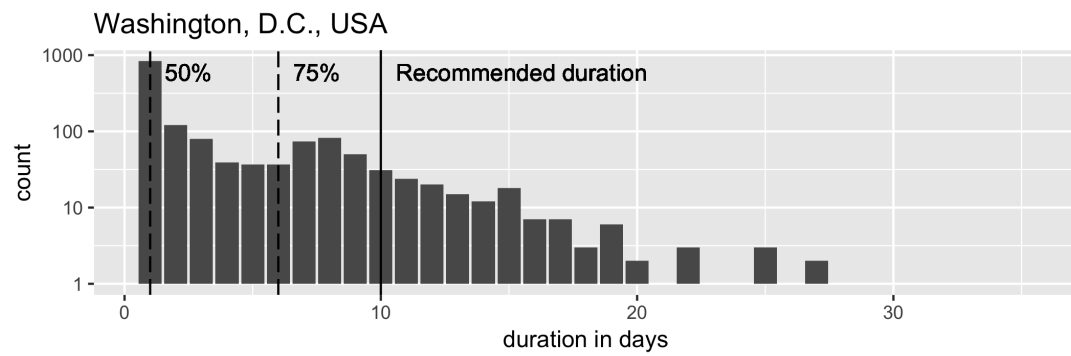

Figure 5: Distribution of the durations of blocks of tourist stays in Washington, D.C., USA

Example. To visualize our approach, we show how the algorithm would calculate the personalized duration of a sample user's visit to Washington, D.C. To that end, we calculate the quantiles of the previously visited cities. In our example, the user made three previous visits, spending 16 days in Tokyo, 10 days in Jakarta, and 7 days in London. We have visualized the distributions of the durations of blocks in the three cities in Figures 2, 3, and 4. The durations of these trips reveal that our user is a relatively slow-paced traveler compared to others, as her lengths of stays are toward the right side of the distributions. The trip to Tokyo was at 97%, the stay in Jakarta at 81%, and the visit to London at 96% of the cumulative distribution function. To aggregate the user's pace over the previous trips, we can calculate the mean percentile, which is 91%. We can then find that percentile in the distribution of visits to Washington, D.C., where the 91th percentile of the distribution is at 10 days (see Figure 5). Therefore, this would be the recommended duration of stay.

CONCLUSIONS

In this paper, we have identified and examined the problem of determining the amount of item consumption in recommender systems. To solve this problem, additional information about the domain and the user's preferences is required. We showcased an approach to determining the personalized duration of a stay in a city, based on the analysis of mobility data from location-based social networks. The underlying method is, however, generalizable to similar problems, given the availability of appropriate data. We argue that such data are indeed often available, especially in commercial recommender systems. In the tourism sector, airlines and hotel portals have a long history of user data and, which they could easily leverage when making recommendations. After all, the proposed approach can be used in any recommender systems domain, where the amount of the recommendation matters and where information about the distribution of the quantity is available for both all users and the particular user of interest.

In the future, we plan to extend our analysis using more trips from different LBSNs and to assess our approach by using offline evaluations that involve cross-validation as well as user studies.

REFERENCES

[1] Munmun De Choudhury, Moran Feldman, Sihem Amer-Yahia, Nadav Golbandi, Ronny Lempel, and Cong Yu. 2010. Automatic Construction of Travel Itineraries Using Social Breadcrumbs. In *21st ACM Conference on Hypertext and Hypermedia (HT '10)*. ACM, New York, NY, USA, 35–44. https://doi.org/10.1145/1810617.1810626

[2] Linus W. Dietz. 2018. Data-Driven Destination Recommender Systems. In *26th Conference on User Modeling, Adaptation and Personalization (UMAP '18)*. ACM, New York, NY, USA. https://doi.org/10.1145/3209219.3213591

[3] Linus W. Dietz, Daniel Herzog, and Wolfgang Wörndl. 2018. Deriving Tourist Mobility Patterns from Check-in Data. In *WSDM Workshop on Learning from User Interactions*. Los Angeles, CA, USA.

[4] Linus W. Dietz, Rinita Roy, and Wolfgang Wörndl. 2019. Characterisation of Traveller Types Using Check-in Data from Location-Based Social Networks. In *Information and Communication Technologies in Tourism*, Juho Pesonen and Julia Neidhardt (Eds.). Springer, Cham, 15–26.

[5] Daniel Herzog and Wolfgang Wörndl. 2014. A Travel Recommender System for Combining Multiple Travel Regions to a Composite Trip. In *CBRecSys@RecSys*. Foster City, CA, USA, 42–48.

[6] Qi Liu, Yong Ge, Zhongmou Li, Enhong Chen, and Hui Xiong. 2011. Personalized Travel Package Recommendation. In *IEEE 11th International Conference on Data Mining (ICDM '11)*. IEEE, Vancouver, BC, Canada, 407–416. https://doi.org/10.1109/icdm.2011.118

[7] Joan Melià-Seguí, Rui Zhang, Eugene Bart, Bob Price, and Oliver Brdiczka. 2012. Activity Duration Analysis for Context-aware Services Using Foursquare Check-ins. In *International Workshop on Self-aware Internet of Things (Self-IoT '12)*. ACM, New York, NY, USA, 13–18. https://doi.org/10.1145/2378023.2378027

[8] Min Xie, Laks V. S. Lakshmanan, and Peter T. Wood. 2012. Composite recommendations: from items to packages. *Frontiers of Computer Science* 6, 3 (June 2012), 264–277. https://doi.org/10.1007/s11704-012-2014-1

[9] Dingqi Yang, Daqing Zhang, Longbiao Chen, and Bingqing Qu. 2015. NationTelescope: Monitoring and visualizing large-scale collective behavior in LBSNs. *Journal of Network and Computer Applications* 55 (Sept. 2015), 170–180. https://doi.org/10.1016/j.jnca.2015.05.010