Travelers vs. Locals: The Effect of Cluster Analysis in Point-of-Interest Recommendation

Pablo Sánchez[1,*,†], Linus W. Dietz[2,**,‡]

[1] Universidad Autónoma de Madrid. 28049, Madrid, Spain
[2] Technical University of Munich. 85748, Garching, Germany

Abstract

The involvement of geographic information differentiates point-of-interest recommendation from traditional product recommendation. This geographic influence is usually manifested in the effect of users tending toward visiting nearby locations, but further mobility patterns can be used to model different groups of users. In this study, we characterize the check-in behavior of local and traveling users in a global Foursquare check-in data set. Based on the features that capture the mobility and preferences of the users, we obtain representative groups of travelers and locals through an independent cluster analysis. Interestingly, for locals, the mobility features analyzed in this work seem to aggravate the cluster quality, whereas these signals are fundamental in defining the traveler clusters. To measure the effect of such a cluster analysis when categorizing users, we compare the performance of a set of recommendation algorithms, first on all users together, and then on each user group separately in terms of ranking accuracy, novelty, and diversity. Our results on the Foursquare data set of 139,270 users in five cities show that locals, despite being the most numerous groups of users, tend to obtain lower values than the travelers in terms of ranking accuracy while these locals also seem to receive more novel and diverse POI recommendations. For travelers, we observe the advantages of popularity-based recommendation algorithms in terms of ranking accuracy, by recommending venues related to transportation and large commercial establishments. However, there are huge differences in the respective travelers groups, especially between predominantly domestic and international travelers. Due to the large influence of mobility on the recommendations, this article underlines the importance of analyzing user groups differently when making and evaluating personalized point-of-interest recommendations.

Keywords

Point-of-Interest recommendation, User Modeling, Human mobility analysis, Offline evaluation,

1. Introduction

Recommender systems are prevalent in numerous areas including videos or movies (Netflix, Youtube), books (Goodreads), consumer products (Amazon), or social contact recommendations (Twitter, LinkedIn) [1]. In the travel and tourism domain, point-of-interest (POI) recommendation is an interesting challenge, where the items to be recommended are venues to be visited when the users arrive at a specific city or region [2, 3]. To perform POI recommendations, much of the data available to the scientific community stems from location-based social networks (LBSNs), such as Foursquare, Gowalla, or Yelp [4, 5]. LBSNs are so frequently used in research because the data usually comprises of several countries and provides additional information

Barcelona '22: 30th Conference on User Modeling, Adaptation and Personalization (UMAP 2022). July 04–07, 2022
pablo.sanchezp@uam.es (P. Sánchez); linus.dietz@tum.de (L. W. Dietz)
0000-0003-1792-1706 (P. Sánchez); 0000-0001-6747-3898 (L. W. Dietz)

about social interactions between the users. Despite the richness and availability of LBSN data, POI recommendation has specific aspects that differ from the conventional recommendation of movies, books, or music that affect the recommenders' performance, including, the implicit information and repeated interactions, as users may check into at the same venue more than once; the relevance of external influences, such as social, temporal, sequential, and, most importantly, the geographical influence, since users tend to visit nearby locations [6, 7, 5]. Finally, the sparsity of the interaction data is typically more severe: For example, the global check-in Foursquare data set from [8] has a density of 0.0034%, making recommendations more difficult than the traditional scenario, such as the well-known Movielens25M data set[1] with a density of 0.2489%.

In addition to the abovementioned issues that affect the performance of the recommenders, we must also consider the different types of users that can be found in LBSNs. Traditionally, when measuring the recommendation quality in offline settings, all users are treated in the same way; hence, there is hardly a recommender systems study that does not report accuracy metrics for each algorithm such as Precision or nDCG averaged over all users, although the focus of evaluation has shifted from only accuracy to further measures, such as novelty, diversity, or serendipity [9, 10]. Recently, researchers have pointed to the importance of analyzing the characteristics of different types of users, e.g., based on their age, gender, and cultural diversity, to detect a possible bias toward certain users in the models [11, 12, 13].

Considering these issues, we analyze to what extent the performance of POI recommendation algorithms differs among different user clusters obtained by analyzing various features. For this, we use a set of well-known cities namely Istanbul, Mexico City, Tokyo, New York, and London, and separate the users into locals and travelers based on them being in their home city or on a visit. For discovering groups within these two categories, we characterize the two user groups based on the behavior they exhibit using various features, thereby focusing on mobility patterns and the types of the visited venues. In the cluster analysis, we obtain different user clusters which we use to analyze the performance of different recommendation algorithms in each of the obtained subclusters in terms of ranking accuracy, novelty, and diversity.

The structure of this paper is as follows: After positioning our approach within literature in Section 2, we describe the process to compute the behavioral metrics and to obtain the different user groups according to the check-ins they performed in Section 3. In Section 4, we explain the experimental procedure followed in the experiments and describe the results obtained in Section 5. Finally, we present our conclusions and future research directions.

2. Related Work

In the tourism domain, there is a considerable variety of categorizing the behavior of many types of travelers visiting a particular region. Such types of travelers have been identified using various methods, such as factor analyses or clustering [14, 15]. For example, tourists have been categorized based on their cultural motives and their cultural depth experience [16], while Yiannakis and Gibson used a three-dimensional scaling analysis between familiarity-strangeness, stimulation-tranquility, and structure-independence to identify 13 different touristic

[1] Movielens25M data set: https://grouplens.org/datasets/movielens/

roles [17]. A more recent article by Neidhardt et al. developed the “Seven Factor Model” wherein the tourist profiles were derived from seven basic factors in which the score of each factor was determined by a set of images selected by the user whose factor score was previously decided by experts [18]. These approaches, thus, established frameworks for categorizing tourists, however, the identified categories are based on a different data source to the domain of the actual recommender system. When developing a new tourism recommender system, one would need to find mappings for both the items and the users to be able to utilize such categorizations. Hence, in this study we would like to determine whether it is possible to obtain different user groups by applying clustering techniques on the same data that is also used in the recommender system. For this, we analyze the user behavior in a Foursquare data set [8], discover groups using cluster analysis and then train and evaluate POI recommenders on the same data set to detect if there are major differences in the recommendations produced to these groups in terms of relevance, novelty and diversity. We are aware that there are other POI recommendation works that apply clustering, like [19, 20, 21]. However, in those articles, the researchers used these techniques to find user groups with common behavior to generate recommendations, while in our work, we identify these user groups based on whether they exhibit a more traveler or local behavior and detect if there are substantial differences in the recommendations received by them.

This article extends and combines two previous studies: In the first one [22], we established trip mining algorithms for LBSN data and already used the global Foursquare check-in data set [8] to identify four different trip types based on trip trajectories. In this work, however, our focus was solely on travelers, not considering the mobility of users while being at their home cities. The other study [23] analyzed the needs of different user types in POI recommendations, by categorizing Foursquare users into different cities into tourists and locals and analyzing the performance of the recommenders in both locals and foreigners. However, as there are many different types of users within these groups [22, 17, 18], we refine this initial analysis by investigating the performance of different recommendation algorithms in each of the user groups in detail. For this, we perform two independent cluster analyses within the travelers and locals, which is driven by the behavior of the users on the global check-in Foursquare data set.

3. User Behavior Characterization and Cluster Analysis

In this first step, we aim to find coherent groups of users that can be discriminated based on information that is relevant to POI recommendation and can be extracted from LBSNs. When performing cluster analysis, the features selected shape the outcome, so it is imperative to compute features that actually help to define the user characteristics. Using a global-scale check-in data set from Foursquare[2] made public by the authors [8], we aim to determine expressive features to characterize different sub-groups within two distinct classes of users: travelers and locals. This separation of travelers and locals is necessary, because the behavior on LBSNs differs significantly depending on the user being home at a city or if she is on a visit. Consequently, there are different features to capture the user behavior.

[2]Foursquare: global-scale check-in data set: https://sites.google.com/site/yangdingqi/home/foursquare-dataset

3.1. Data Preprocessing

This Foursquare data set contains a total of 33M check-ins from 415 different cities globally. Starting from the complete data set, we performed the following preprocessing steps to eliminate noise and ensure a higher data quality: We first removed users with consecutive check-ins of less than 60s, as well as consecutive check-ins in the same POI and check-ins with an unrealistic transition speed of more than 343 m/s. Next, we enforced 10-core for users and POIs, i.e., removed interactions so that ultimately all remaining users have at least 10 interactions and each POI has at least 10 visits. Finally, we split the processed data set following a temporal partition in which 80% of the most ancient interactions are sent to the training set, whereas the other 20% is used as the test set.

Using the information in the training set, we performed two cluster analyses, independently for locals and travelers. To perform this study correctly, it is essential to know a user's home because only check-ins of the home city of a user should be used to compute their behavior as a local; likewise, a user's travel behavior should solely be characterized using check-ins outside of the home city. For determining a user's home in the context of LBSN check-in data, several methods exist [24]; however, taking the city where most check-ins are done consistently produces highly accurate results when used along with a threshold. As such, we determined exactly one city for each user as home city using a threshold of at least 50% of check-ins needed to be performed in the most frequent city. This step excludes another 8,548 (6.20%) users with an unclear home from the training data, resulting in 129,294 valid users in the training set.

3.2. Local Behavior Cluster Analysis

To discover distinct groups of user activity in their home town, we exclusively analyzed check-ins they have performed in their home cities and computed various features including mobility metrics, such as the radius of gyration, the mean distance from the city center, and the mean distance between consecutive check-ins. Further features describe the activity of the users, e.g., the mean time between check-ins, the activity period, the number of check-ins, and the number of unique POIs visited. Finally, we also count check-ins in relevant categories separately, such as visiting POIs labeled with "Arts & Entertainment," "Outdoors & Recreation," "Food," "Nightlife Spot," and "Shops & Services."

First, we analyze correlations between features and eliminate those that have a high correlation > 0.7, as they are redundant. Likewise, we eliminate features that are orthogonal to all other ones identified by very low correlations with other attributes $[-0.1; 0.1]$. These features are essentially treated as noise by the clustering algorithm and, thus, decrease the quality of the discovered groups. Concretely, this step resulted in the elimination of the following metrics: mean check-ins per day, total number of check-ins, and the number of check-ins in "Colleges & Universities."

Using the k-means algorithm, we systematically analyzed the outcome of the algorithm using the Euclidean Distance and min-max normalized features. Examining the quality of the resulting clusters using different values for k, we observed that the segmentation quality was very low, despite having performed the relevant steps of the prior correlation analysis. Experimenting with different feature combinations, the silhouette width ranged in the area of 0.3 for 3–4 clusters and further dropped with a higher k. However, when dropping the mobility features (radius of gyration, mean check-in distance, and mean distance to city center), we obtained clearly better results, and finally choose the optimal configuration of a silhouette width of 0.57 for $k = 3$. We plot the silhouette width against k in Figure 1a and refer to the details of the final result in Table 1.

Table 1
Cluster results of the 129,294 locals. In the absence of mobility features, the segmentation is mostly driven by the activity level of the users. Values represent the mean/standard deviation.

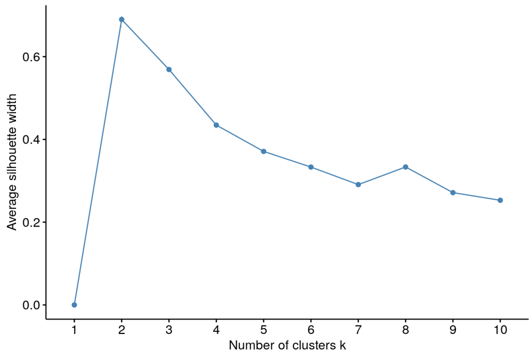
	L1	L2	L3
Name	Low	Medium	High
Ratio	25.3%	28.0%	46.6%
Activity Duration	79.74/40.47	205.65/38.98	341.86/30.75
Unique POIs	14.36/ 9.89	20.63/12.68	26.03/16.66
Arts & Entertainment	1.30/2.70	2.11/3.49	3.58/5.54
Outdoors & Recreation	4.18/ 7.43	5.87/10.01	6.55/12.23
Food	6.65/ 9.03	10.45/12.40	13.67/18.48
Nightlife Spot	1.42/3.60	2.38/4.83	3.73/7.38
Shops & Service	4.43/ 6.27	6.43/ 8.01	8.74/11.61

There are three clusters, two which respectively make up about a quarter of the users and one larger one, containing the remaining 46.6% of the locals. We interpret the fact that the mobility features, such as the radius of gyration and the distance to the city center prevented the algorithm from finding an acceptable segmentation of the locals, as a clear indication that these features are unsuitable for distinguishing different resident groups in the data set at hand. This may be due to several reasons: residents might be more active in their respective districts making it hard to characterize their behavior with metrics in relation to the entire city. In addition, commuting introduces noise, which is difficult to eliminate given the volatile usage of LBSNs during leisure and work time. Finally, the mobility metrics to characterize residents of five different cities might need more careful deliberation: cultural and geographic circumstances could be too different to find universal clusters across all cities. This means that the clustering result of the locals is mostly influenced by the user activity level.

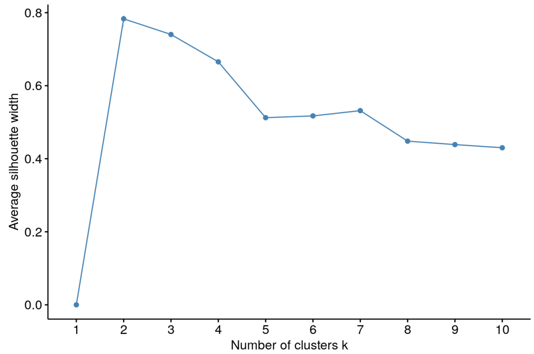
3.3. Traveler Behavior Cluster Analysis

Similar to the locals, we analyzed the behavior of the users when traveling outside of their respective home cities. The processing was performed using the `tripmining` library[3], which segments the user's check-ins into periods of being at home and in other cities [22]. Consecutive periods abroad are regarded as trips, provided certain data quality criteria are met. Unlike the analysis of the local behavior, these quality criteria are necessary because we need to know the location of the user at any time. However, the nature of check-in-based data is that we only know the user location when she used the Foursquare app, thus, we have an incomplete view

[3]`tripmining` library: https://github.com/LinusDietz/tripmining



(a) Locals: $k = 3$ was chosen as final result, since this was the last value above 0.5 and with larger values for k, the silhouette width only decreases gradually.



(b) Travelers: We chose $k = 4$, since with higher k, the silhouette width plateaus.

Figure 1: Determining the number of clusters using the silhouette width.

of the periods between the check-ins. This uncertainty is acceptable in a global travel scenario, since users typically only travel to a few cities per day and it is possible to quantify the date quality using various metrics. In this case, we used the default settings of the tripmining library: A minimum check-in density of 0.5, which means that there is on average at least one check-in in two days during the trip, a minimum duration of two calendar days between the first and the last check-ins of a trip, and a maximum of three days without any check-in. These metrics limit the uncertainty involved when working with incomplete information, which is inherent to check-in based data. As a result, we obtained 38,903 travelers who did a total of 64,316 trips.

We aggregated all trips of a traveler as their traveler profile, and again used the same method used for the locals to select the features. The number of stays in cities (non-distinct) and the number of "Food" check-ins were eliminated due to a high correlation to the number of trips. The final features lead to four clusters with a silhouette width of 0.68. We chose $K = 4$ as the optimal number of clusters, as the silhouette width was just slightly lower than $K = 3$ (0.73), but clearly higher for $K \geq 5$, which was around 0.5 (cf. Figure 1b).

The four traveler clusters tabulated in Table 2 show similar groups as the clustering of Dietz et al. [22], although their work clustered trips, whereas we aggregated the trip metrics per traveler before clustering. With around 81% of the traveling users, T3 (Domestic) is the most numerous group comprising travelers whose trips were almost exclusively domestic close to their home cities. T1 (Foreign Cities) are infrequent travelers with only 1.33 trips that are mainly international, where the users were quite stationary at their destination, as can be seen in the low radius of gyration. T4 (Globetrotters) is similar; however, this group of intercontinental travelers, was more into POIs of the "Arts & Entertainment" category than T1. The high radius of gyration in Globetrotters can be an artifact of airfare stopovers because such check-ins are also included in the trips. Finally, T2 (Active Vacationers) is also a small cluster, but it has the most active travelers with 2.77 trips visiting many unique cities both in their own country and abroad.

Table 2
Cluster results of the 38,903 travelers. The discovered groups shed light on the preferred type of trips the users did. Values are the mean/standard deviation.

	T1	T2	T3	T4
Name	Foreign Cities	Active Vacationers	Domestic	Globetrotters
Ratio	11.1%	5.0%	80.6%	3.2%
Ratio Domestic Trips	0.09%	55.36%	99.94%	0.46%
Displacement	1324.58/1086.56	1746.56/1806.27	503.19/ 673.21	7599.87/2715.97
Radius of Gyration	263.34/ 556.04	1783.14/2029.40	108.96/ 285.24	1968.35/2818.53
Number of Trips	1.33/1.01	2.77/1.20	1.64/1.44	1.30/0.89
Unique Cities	1.64/0.96	3.45/1.54	1.58/0.94	2.30/1.44
Arts & Entertainment	0.44/0.85	0.89/1.29	0.38/0.81	0.72/1.35
Outdoors & Recreation	0.75/1.33	1.53/2.18	0.95/1.92	0.80/2.31
Nightlife Spot	0.27/0.76	0.68/1.25	0.44/1.12	0.37/1.93
Shops & Service	0.90/1.62	1.65/2.22	0.98/1.80	0.90/1.62

3.4. Summary

We characterized Foursquare users by features that can be computed exclusively from analyzing their check-ins. The independent characterization of the users' check-in behavior in their home city and during travel allowed us to discover three and four distinct groups for locals and travelers, respectively. Our main takeaway from the cluster analysis is that the mobility metrics explored in our work seem to be unsuitable for characterizing locals in our LBSN check-in data set, as the clustering algorithms struggle to find distinct groups using these features. This also implies that if we mix travelers and locals users when evaluating POI recommendation algorithms, we will likely observe disparate results due to the fact that we may not adapt well to the interests of any of them, as – quite unsurprisingly – their behavior differ considerably. We use these groups to systematically investigate the effect of using such cluster information of users on the performance of POI recommender systems.

4. Experimental Settings

Once we establish different user groups applying the clustering, we now describe the setup followed for performing POI recommendations. The global-scale check-in data set from Foursquare comprises a total of 33M check-ins from over 415 different cities around the world. Starting from the complete data set, we performed the preprocessing steps and the temporal split, as stated in Section 3. With the processed data set, for producing the recommendations, we decided to select a set of large metropolises from around the world with different densities: Istanbul, Mexico City, Tokyo, New York, and London. We decided to work with these cities independently (training and testing the recommenders separately for each city) because as the geographical information is exploited by many POI recommendation approaches, it may be counterproductive to mix check-ins from geographically distant regions. In Table 3, we present the statistics of the different cities, showing the number of total users users, venues, check-ins, unique check-ins,

Table 3
Statistics of the data set and cities used in the experiments. $|\mathbf{U}|$, $|\mathbf{V}|$, $|\mathbf{C}|$, and $\frac{|\mathbf{C}|}{|\mathbf{U}|\cdot|\mathbf{V}|}\%$ represent the number of users, venues, check-ins, and the density, respectively. As in LBSNs, some users may check-in in the same venue more than once, we also report in column $|\mathbf{C}|_u$ the number of unique check-ins and $\frac{|\mathbf{C}|_u}{|\mathbf{U}|\cdot|\mathbf{V}|}\%$ represents the density with the unique check-ins.

City	Split	|U|	|V|	|C|	$\|\mathbf{C}\|_u$	$\frac{\|\mathbf{C}\|}{\|\mathbf{U}\|\cdot\|\mathbf{V}\|}\%$	$\frac{\|\mathbf{C}\|_u}{\|\mathbf{U}\|\cdot\|\mathbf{V}\|}\%$
Filtered data set	Full	139,270	251,115	9,266,149	4,354,336	0.02650	0.01245
	Training	137,842	248,692	7,412,919	3,596,596	0.02162	0.01049
	Test	108,213	196,945	1,853,230	1,134,909	0.00870	0.00532
Istanbul	Full	29,307	20,366	1,569,015	821,683	0.26288	0.13767
	Training	26,894	19,976	1,189,646	645,536	0.22144	0.12016
	Test	21,780	17,226	379,369	248,157	0.10112	0.06614
Mexico City	Full	5,944	7,978	286,638	147,850	0.60445	0.31178
	Training	5,690	7,948	237,188	125,675	0.52447	0.27789
	Test	4,018	6,442	49,450	32,616	0.19104	0.12601
Tokyo	Full	6,631	5,543	227,391	122,814	0.61866	0.33414
	Training	6,213	5,534	186,248	103,768	0.54169	0.30180
	Test	4,194	4,831	41,143	28,211	0.20306	0.13924
New York	Full	8,170	3,557	109,611	68,988	0.37718	0.23739
	Training	7,238	3,548	92,790	59,342	0.36133	0.23108
	Test	3,319	2,867	16,821	12,728	0.17677	0.13376
London	Full	4,235	1,612	43,794	26,472	0.64150	0.38776
	Training	3,520	1,607	35,516	21,697	0.62786	0.38357
	Test	1,749	1,361	8,278	6,108	0.34776	0.25660

and both the training and test sets used in each independent city. Note that the filtered dataset was used to generate the locals and travelers clusters. Notably, this table includes all users found in each city, even those whose home towns are unclear (and hence no traveler nor local cluster associated). Because we performed a temporal split, there might be new users in the test set with no user cluster associated, as the cluster analysis was performed solely on the training set. Analyzing the values in this table, we want to highlight some relevant observations before showing the actual experimental results. First, from Table 3, the check-in repetitions represent a relevant percentage of the interactions (the percentage of unique check-ins reach at most 60%), making it difficult to recommend new POIs to users. Further, only a total of 38,903 users were observed to be traveling in the training period, providing the algorithms less training data than the locals.

4.1. Algorithms

In this section, we briefly list the algorithms used in our experiments, which can be categorized into classic and POI recommendation algorithms (the parameters tested in the recommenders are shown in Table 4). For their exact formulations, we refer the reader to the respective references.

- Classic recommendation algorithms:
  - Rnd: performs recommendations of venues randomly.
  - Pop: recommends to the target user the venues that have been visited by the largest number of users.
  - UB/IB: non-normalized user and item-based neighborhood approaches [25, 26].
  - HKV: matrix factorization (MF) algorithm that uses Alternate Least Squares for optimization (from [27]).
  - BPRMF: Bayesian Personalized Ranking (a pairwise personalized ranking loss optimization algorithm) using a MF approach (from [28]). We used the version from the MyMedialite[4] library.
- Specific algorithms for POI recommendation:
  - IRENMF: weighted MF method from [29]. This method incorporates geographical information in two different ways: instance level influence (users tend to visit neighboring locations) and region-level influence (they assume that the user preferences are shared in the same geographical region).
  - GeoBPR: geographical BPR. POI recommender optimized using BPR [30]. It analyzes the POIs visited by the target user and assumes that she will prefer to visit new POIs that are close to the ones she visited previously.
  - FMFMGM: probabilistic MF with multi-center Gaussian model. It is an hybrid approach proposed by [31] that combines Probabilistic MF (PMF) with a Multi-center Gaussian Model technique (MGM).
  - RankGeo-FM: a ranking-based MF model proposed in [6]. They model the geographical influence by exploiting the geographical neighbors POIs with respect to the target POI using an additional latent matrix for the users.
  - PGN: popularity, geographical, and user-based neighborhood. Hybrid approach that combines the popularity algorithm (Pop), user-based neighborhood (UB), and a geographical recommender that recommends to the target user the venues closer to the average geographical position of all the venues visited by the user. The final score is an aggregation of every item score provided by each recommender after normalizing its values by the maximum score of each method.

4.2. Experimental Setup

As we mentioned in Section 4, we applied a temporal split in which we selected the 80% of the most ancient interactions of the filtered data set as the training set and the rest as the test set. Afterward, we selected the check-ins for each city and trained the recommenders using the data of each city independently, as done in many state-of-the-art POI recommendation studies [32, 29, 6, 30], where the authors test their approaches in a subset of cities or regions. We followed the "TrainItems" methodology [33], in which we consider for each user u all venues of the training set that have not been visited by u. We firmly believe that this approach is suitable because as opposed to repeated consumption of items, in e.g., the music domain, the inherent

[4]MyMedialite library: http://www.mymedialite.net/

Table 4
Parameters tested in the recommenders. The best configurations are selected by maximizing nDCG@5.

Rec	Parameters
UB/IB/PGN	Sim = {Vector Cosine, Set Jaccard}, $k = \{20, 40, 60, 80, 100, 120\}$
HKV	Iter = 20, Factors = $\{10, 50, 100\}$, $\lambda = \{0.1, 1, 10\}$, $\alpha = \{0.1, 1, 10, 100\}$
BPRMF	Factors = $\{10, 50, 100\}$, BiasReg = $\{0, 0.5, 1\}$, LearnRate = 0.05, Iter = 50, RegU = RegI = $\{0.0025, 0.001, 0.005, 0.01, 0.1\}$, RegJ = RegU/10
IRENMF	Factors = $\{50, 100\}$, geo-$\alpha = \{0.4, 0.6\}$, $\lambda_3 = \{0.1, 1\}$, clusters = {5, 50}
FMFMGM	Factors = $\{50, 100\}$, $\alpha = \{0.2, 0.4\}$, $\theta = \{0.02, 0.1\}$, dist = 15, iter = 30, $\alpha_2 = \{20, 40\}$, $\beta = 0.2$, sigmoid = False, LearnRate = 0.0001
RankGeo-FM	Factors = $\{50, 100\}$, $\alpha = \{0.1, 0.2\}$, c = 1, $\epsilon = 0.3$, neighs = $\{10, 50, 100, 200\}$ iter = 120, decay = 1, boldDriver = True, learnRate = 0.001

Table 5
Results of the recommenders for Istanbul, Mexico City, and Tokyo. In bold, we show the highest value for each city in each classic and POI types of recommenders in each metric. In bold with a dagger, we show the highest values in each city.

		Istanbul				Mexico City				Tokyo			
Type	**Rec**	**nDCG**	**EPC**	**IC**	**UC**	**nDCG**	**EPC**	**IC**	**UC**	**nDCG**	**EPC**	**IC**	**UC**
Classic	Rnd	0.000	†**0.995**	†**19,886**	†**21,780**	0.000	†**0.988**	†**7,286**	†**4,018**	0.000	†**0.990**	†**5,422**	†**4,194**
	Pop	0.033	0.129	25	†**21,780**	†**0.051**	0.291	14	†**4,018**	†**0.051**	0.274	19	†**4,194**
	UB	**0.040**	0.537	2,491	19,279	0.026	0.693	2,308	3,764	0.041	0.439	855	3,776
	IB	0.036	0.605	9,247	19,362	0.019	0.842	4,765	3,764	0.037	0.633	3,151	3,776
	BPRMF	0.036	0.568	3,154	19,367	0.038	0.331	156	3,764	0.044	0.338	414	3,776
	HKV	0.025	0.713	950	19,367	0.018	0.820	704	3,764	0.028	0.576	78	3,776
POI	IRENMF	0.043	0.541	1,243	19,367	0.034	0.635	923	3,764	0.043	0.519	1,220	3,776
	GeoBPR	0.042	**0.626**	722	19,367	0.041	0.421	196	3,764	0.046	0.403	300	3,776
	FMFMGM	0.028	0.356	259	19,367	0.019	0.591	300	3,764	0.039	0.375	117	3,776
	RankGeo-FM	0.039	0.567	2,324	19,367	0.022	**0.673**	1,578	3,764	0.033	**0.593**	**1,870**	3,776
	PGN	†**0.044**	0.228	**3,032**	†**21,780**	**0.051**	0.435	**2,242**	†**4,018**	**0.050**	0.377	1,559	†**4,194**

value of POI recommendation is to suggest new places for users to be discovered. Finally, as we mentioned above, we will not only measure the performance of the recommendations in terms of nDCG, but also we will analyze the novelty (in terms of EPC), the diversity (in terms of Aggregate Diversity, or Item Coverage, IC), and the user coverage (UC) of the different algorithms. Unless stated otherwise, the results of all metrics are shown @5. The novelty and diversity metrics are defined as:

- Expected Popularity Complement (EPC): a novelty metric that gives a higher value (and hence, more novel) to those items that are less popular [34]. It is formulated as: $1/|R_u| \sum_{i \in R_u} (1 - |\mathcal{U}_i|/|\mathcal{U}_{tr}|)$, where R_u denotes the recommendation list of a user, $\mathcal{U}_{tr}$ represents the set of users in the training set, and $\mathcal{U}_i$ is the set of users who rated item i in the training set. However, in this study, we will show a normalized EPC value by applying the min-max normalization.
- IC (Item Coverage, also referred to as Aggregate Diversity) diversity metric that measures the number of different items that an algorithm is able to recommend [35]. It is formulated as $|\bigcup_{u \in \mathcal{U}_{rec}} R_u|$, where $\mathcal{I}_{tr}$ denotes the set of items in the training set and $\mathcal{U}_{rec}$ represents the set of users to whom we have provided recommendations.
- UC (User Coverage): measures the number of users that the recommender is able to provide recommendations. It is formulated as $|\mathcal{U}_{rec}|$.

Table 6
Results of the recommenders for New York and London. The same configuration as in Table 5.

Type	Rec	New York nDCG	EPC	IC	UC	London nDCG	EPC	IC	UC
Classic	Rnd	0.000	†**0.991**	†**3,509**	†**3,319**	0.003	†**0.959**	†**1,603**	†**1,749**
	Pop	†**0.114**	0.436	13	†**3,319**	**0.046**	0.168	11	†**1,749**
	UB	0.056	0.688	964	2,317	0.036	0.538	546	1,033
	IB	0.032	0.853	2,407	2,386	0.035	0.766	1,166	1,033
	BPRMF	0.080	0.489	388	2,387	0.039	0.189	11	1,034
	HKV	0.038	0.776	402	2,387	0.015	0.717	88	1,034
POI	IRENMF	0.070	0.617	477	2,387	0.034	0.541	379	1,034
	GeoBPR	0.076	0.502	155	2,387	0.046	0.301	102	1,034
	FMFMGM	0.024	0.683	108	2,387	0.028	0.468	203	1,034
	RankGeo-FM	0.028	**0.773**	**1,892**	2,387	0.020	**0.673**	**1,117**	1,034
	PGN	**0.108**	0.542	1,505	†**3,319**	†**0.050**	0.241	332	†**1,749**

5. Analysis of Results

5.1. Performance of Recommenders in Each City

Before showing the results obtained for each of the user clusters, in Tables 5 and 6, we present the results obtained by the recommenders in the five aforementioned cities. In this case, the value of each metric is computed for every recommended user (represented in the UC metric) and then returning the average, as is the standard in the literature. Analyzing these results, we first note the low values obtained in nDCG. This is due to several causes: the high sparsity of data, the temporal split in which is common to find new relevant venues that cannot be recommended as they do not appear in the training set, and tendency of users checking-in in the same POI repeatedly. As we use the “TrainItems” methodology, those venues are unsuitable to be recommended, as the objective is to recommend new venues to users.

Only the Rnd, Pop, and PGN have complete coverage at the user level because when a temporal split is performed, there are users in the test set that do not appear in the training set. Both Rnd and Pop are not personalized, so they can perform recommendations to new users. Although PGN is a personalized recommender, it will fall back to recommend popular POIs to cold-start users in the test set, but not on the training set. With respect to ranking accuracy, novelty, and diversity, we note that the Pop recommender, which is generally the best in terms of relevance in all cities, except Istanbul, in which the best algorithm is the PGN model obtaining 0.044 in nDCG (showing the strong popularity bias existing in this domain), is the worst in both novelty and diversity. Moreover, PGN always obtains competitive results in ranking accuracy, whereas it obtains higher values in novelty and diversity than Pop (although it is not the best model in any dimension). This illustrates one of the fundamental problems in recommendation, which is that it is nearly impossible to find a model that obtains the best performance in all metrics, making it indispensable to define algorithms that exhibit a balance between the different dimensions being analyzed [35].

Regarding the other recommenders, we can observe that in general, POI recommendation algorithms tend to obtain better results than the classical recommenders, excluding Pop, at least in terms of ranking accuracy. Nevertheless, some classic recommenders, such as the UB are still competitive. Although this shows that classic recommender algorithms are still useful to be considered as baselines, it is a clear indication of the importance of considering the geographical influence in the POI recommendation domain. With respect to these models, we can observe that besides PGN, the IRENMF recommender is one of the best in all dimensions. This result is consistent with previous findings [7, 23], where it obtained a good balance between accuracy, novelty and diversity. Nevertheless, we observe that GeoBPR, in general, outperforms IRENMF in terms of ranking accuracy with the exception of Istanbul.

In the next section, we will perform an in-depth analysis of the performance of the most representative models in different groups of both travelers and locals shown in Sections 3.2 and 3.3.

5.2. Performance of Recommenders in Specific User Groups

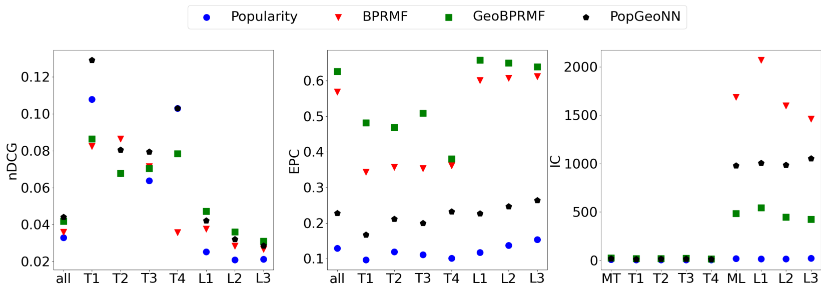


Figure 2: Results for Istanbul.

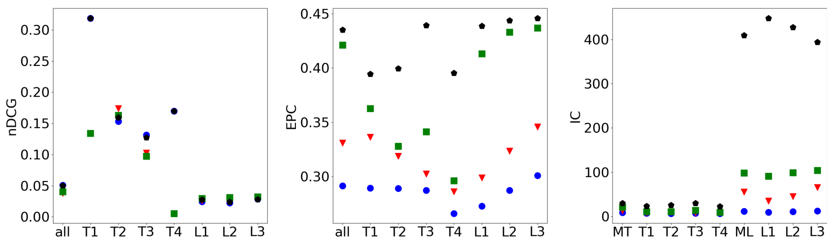


Figure 3: Results for Mexico City.

Having shown the results of the recommenders by computing the average among all the users, we turn our focus on analyzing the value obtained in each metric for the different cluster groups for both locals and travelers. Hence, we show these results in Figures 2, 3, 4, 5, and 6 for the five abovementioned cities. For those figures, we show the performance of the clusters of the travelers (denoted with T1, T2, T3, and T4), the locals (L1, L2, and L3), and all users in the test set (all). We present three metrics in those figures: nDCG (for ranking accuracy), EPC (novelty), and IC (diversity). Regarding this last metric, notably, according to its formulation, as it does not compute the average between the users to whom we have recommended, it may

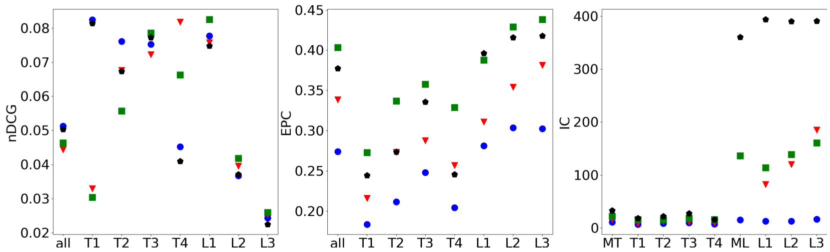


Figure 4: Results for Tokyo.

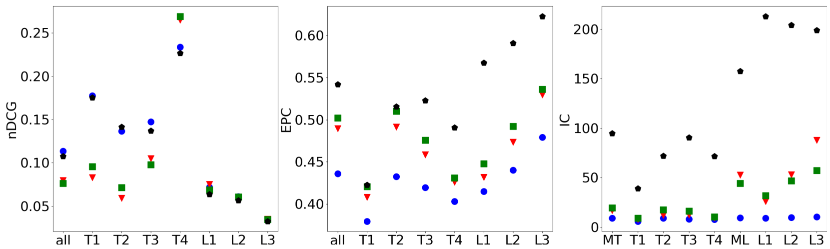


Figure 5: Results for New York City.

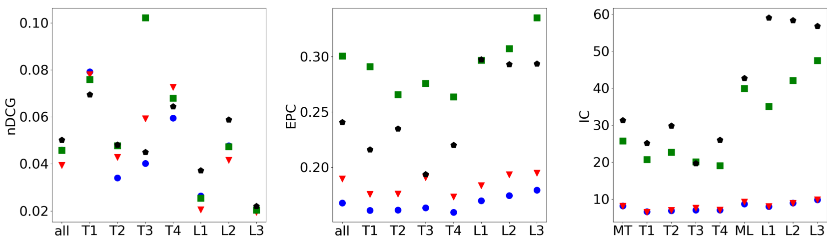


Figure 6: Results for London.

obtain different results in each user cluster depending on the number of users who belong to each group. For example, if we compare the diversity between T3 and T1 using this original formulation, we would obtain a much lower diversity in T1, because the number of travelers in the first cluster is lower than in the third one. To mitigate this lack of normalization, we compute this metric performing different subsamples. Hence, we selected for each major group (travelers and locals) the cluster with the smallest number of users and then computed the value of the IC for this number of random users. The final value is the mean after repeating the sampling $1,000$ times, thereby making the values comparable. This is why instead of the “all” label, we use two additional ones when representing the IC metric, “MT” (mean of travelers) and “ML” (mean of locals), which would be used for selecting a subsample of all users with the size of the lowest traveler and local group, respectively. We repeat this process $1,000$ times, and then the result shown is the average of the $1,000$ runs. Finally, due to the large number of recommenders, we decided to show for each city the performance of Pop and BPRMF from the classical recommenders and GeoBPR and PGN from the POI recommenders, as they are the algorithms which generally obtain the best nDCG results.

Analyzing the figures, we can observe interesting effects. First of all, travelers generally obtain higher values than locals in terms of accuracy in most cities (e.g., in Istanbul and in Mexico City all traveler groups obtain higher values in terms of nDCG than any local cluster), despite being the group with less users (e.g., 7% in the case of Mexico City and 24% of New York users in the test set are travelers). However, notably, travelers generally have a slightly lower novelty than locals, indicating that they tend to receive recommendations of more popular POIs. This makes sense, because when a tourist visits a city, she is more likely to visit the most popular venues than if she is a local. Besides, we analyzed the top-5 most popular venues in each city and we observed that most of them belong to transport and commerce categories. For example, both airports (e.g., Kennedy, Atatürk and Benito Juárez, in New York, Istanbul, and Mexico City respectively) and train stations (Euston in London and Akihabara in Tokyo) are some of the venues that have received most visits in the training sets. In addition, shopping malls and commerce districts like Harrods (London), Perisur (Mexico City), and Times Square (New York) are also in the top-5 most popular venues.

Moreover, we can observe how locals tend to receive more diverse recommendations. This again may be because locals are commonly sightseeing extensively within their home city. Furthermore, locals are more likely to have visited numerous different POIs in the training set (including the most popular ones), which are then unavailable for recommendation in the test set. By contrast, most travelers will visit a city for the first time during the evaluation period; thus, it is more probable that they visit one of the recommended popular POIs, which will result in a decrease in novelty and diversity. Finally, as there are far fewer travelers than locals, it is normal that despite having computed the IC metric using the subsamples, we obtain much lower results for travelers than for locals, making a direct comparison between them impossible.

In general, the big picture of these results tends to support the findings of [23], although we performed a different data preprocessing, splitting methodology, and also a different analysis and characterization of travelers and locals. Hence, in addition to the analysis performed for travelers and locals, it is also interesting to study the behavior of the models among the different types of travelers and locals, i.e., all clusters shown in Tables 1 and 2.

First, regarding the travelers, there is no a common behavior in the different cities. For example, T4 is the group that obtains the highest values in nDCG for New York in all recommenders, whereas, in other cities, such as Istanbul and Mexico City, there are some models which obtain very low values for these users. Regarding novelty and diversity, T1 obtains the worst results in the cities of Tokyo and New York, whereas, in London it is one of the best group in both aspects. Despite these discrepancies among the travelers, we also perceive common behavior, such as T2 and T3 generally obtaining similar results. This may be explained by the features shown in Table 2, where we can observe that these two groups have the highest ratio of domestic trips, whereas T1 and T4 tend to make more abroad travels, visiting more popular POIs as we can observe in the performance in both nDCG and EPC metrics. In fact, except for Mexico City, in the rest of the cities the Pop algorithm achieve higher values in EPC for both T2 and T3.

Regarding the locals, in all cities, except Mexico City, L3 is the cluster that obtains the lowest levels in nDCG, comparing it with all locals and travelers groups whereas, in general, it also obtains higher levels of novelty than the other clusters. Besides, for L3, all models have similar nDCG performance, thus, exhibiting much fewer variations in this group than in any other group. From Table 1, besides L3 being the most numerous, it is also the cluster that, in general, contains the most active users (represented by the "Unique POIs" feature). Hence, it is more probable to recommend these users less popular venues given the probability that they have visited more venues before than the other two groups with a lower activity level, making more difficult to recommend to them both novel and relevant venues.

5.3. Discussion

According to the results obtained, if we segment the Foursquare users into different clusters of travelers and locals, we observe well-differentiated behaviors. As most users have been mostly active only in their home cities, there are fewer users to be analyzed belonging to the traveler groups than locals.

When analyzing the recommendations, we found that travelers tend to get higher values in terms of accuracy and lower values in terms of novelty and diversity. Nevertheless, we also observed that for both travelers and locals, the performance of the recommenders is rather low and sometimes the best performing algorithm in the basic popularity recommender, which confirms the trend observed in [23]. This emphasizes the role of the popularity bias in POI recommendation, although we believe that this bias would be worth analyzing more in-depth for this domain, as it has been done in other traditional recommendation scenarios [36, 37].

A relevant insight of this study is that by assessing the quality of the clustering results, it is imperative to use different features to derive the clusters of travelers and locals. For the travelers, we note that the geographical information was especially relevant, as we found four highly differentiated groups according to the ratio of domestic trips and the geographic displacement. Regarding this, we observed that T2 and T3 tend to make more domestic trips, having comparable results in the evaluation metrics of the recommendations. For locals, we found that that the most important features were regarding the activity level, especially in terms of activity duration and the number of unique POIs visited. In this sense, we observed that L3, which was the most numerous, exhibits the highest values in the abovementioned features, whereas it also obtains the lowest performance in terms of ranking accuracy.

Possibly most importantly, with this analysis using an LBSN data set, we showed that different user groups exhibit very different behavior; therefore, it would be misleading to measure the performance of recommendation algorithms for all users as a whole. Especially, when the recommendations should be tailored to specific groups, a "one-size-fits-it-all" algorithm, which seemingly produces good recommendations, might fail for a specific user group. Concretely, we could measure differences of $> 400\%$ between different user groups in terms of nDCG, such as in London and New York using GeoBPR recommender or Mexico City using the PGN algorithm. Besides, we also observed some important differences between user groups when measuring EPC (although generally smaller than in nDCG), in the case of Tokyo for the PGN and BPRMF algorithms. Further, in view of the analyses and results obtained, we would like to raise concerns that this Foursquare data set may not be appropriate to be used in the tourism domain because the vast majority of the users have barely checked-in in more than one city, cf. Tables 1, and 2. However, although this data may be ill-suited for obtaining general conclusions about the mobility patterns of travelers in a real-world environment, we do believe that LBSN data might help tourism applications to recommend novel and diverse venues to users exploiting the interactions of locals, as they will have more knowledge about the interesting venues in a city [38].

6. Conclusions & Future Work

In this paper, we have presented a study on the POI recommendation by classifying Foursquare users into different groups of travelers and locals. To obtain these groups, we analyzed different mobility features for travelers derived from the trips they have done during the observation period. For locals, we observed that the geographical information used in this work is not helpful in computing the different clusters, so we decided to use the information related to the different types of POIs visited by users and the activity level they exhibited. Besides, we analyzed the performance of a wide range of classic and POI-specific recommendation models in the abovementioned travelers and locals clusters in terms of ranking accuracy, novelty and diversity. Regarding the results obtained, we have observed that this Foursquare data set is mostly formed of users who are local to a given city, meaning that this type of data may less-suited for analyzing tourism patterns travelers. However, we verified that despite less available training data for travelers, it is easier to recommend them relevant venues compared to the locals, which we attribute to travelers being more impacted impacted by popularity bias, represented by venues related to transportation and with shops & services. Moreover, we have also observed performance differences among the discovered traveler and local subclusters. Thus, regarding the locals, we have detected that it is more difficult to produce relevant recommendations to the users who have spent much time in their home city. Similarly, recommendations are generally easier to compute for international travelers than for domestic ones, despite most travel being domestic.

Generally, this study strengthens the conclusions of some previous studies [39, 23], but at the same time shows that POI recommendation using LBSN data is more intricate than how many approaches tackle the problem. Different user groups have different needs, which need to be considered by the recommendation algorithms. As future work, we argue that it would be essential to extend this analysis to other LBSNs, such as Gowalla[5] or Brightkite[6], to see if we can obtain similar user groups to the ones discovered in our work. Other data sources might exhibit other features to characterize the users, which raises the research question of which features can be regarded as universal between multiple LBSNs. Finally, we believe that it might be useful to analyze additional features to create the clusters, including geographical, temporal (e.g., add more temporal and geographical restrictions to derive the home cities of the users) and/or the POI categories visited by each of the different types of users, to detect additional biases in the recommendations.

Acknowledgments

This work has been funded by the Ministerio de Ciencia e Innovación reference (PID2019-108965GB-I00) and by the European Social Fund (ESF), within the 2017 call for predoctoral

[5]Gowalla data set: https://snap.stanford.edu/data/loc-gowalla.html
[6]Brightkite data set: https://snap.stanford.edu/data/loc-brightkite.html

contracts.

References

[1] F. Ricci, L. Rokach, B. Shapira, Recommender systems: Introduction and challenges, in: Recommender Systems Handbook, Springer, 2015, pp. 1–34. doi:10.1007/978-1-4899-7637-6_1.

[2] J. Borràs, A. Moreno, A. Valls, Intelligent tourism recommender systems: A survey, Expert Systems with Applications 41 (2014) 7370–7389. doi:10.1016/j.eswa.2014.06.007.

[3] C. Trattner, A. Oberegger, L. B. Marinho, D. Parra, Investigating the utility of the weather context for point of interest recommendations, Information Technology & Tourism 19 (2018) 117–150. doi:10.1007/s40558-017-0100-9.

[4] J. Bao, Y. Zheng, D. Wilkie, M. F. Mokbel, Recommendations in location-based social networks: a survey, GeoInformatica 19 (2015) 525–565. doi:10.1007/s10707-014-0220-8.

[5] P. Sánchez, A. Bellogín, Point-of-interest recommender systems based on location-based social networks: A survey from an experimental perspective, ACM Computing Surveys (2022). doi:10.1145/3510409.

[6] X. Li, G. Cong, X. Li, T. N. Pham, S. Krishnaswamy, Rank-geofm: A ranking based geographical factorization method for point of interest recommendation, in: SIGIR, ACM, 2015, pp. 433–442. doi:10.1145/2766462.2767722.

[7] Y. Liu, T. Pham, G. Cong, Q. Yuan, An experimental evaluation of point-of-interest recommendation in location-based social networks, Proceedings of the VLDB Endowment 10 (2017) 1010–1021. doi:10.14778/3115404.3115407.

[8] D. Yang, D. Zhang, B. Qu, Participatory cultural mapping based on collective behavior data in location-based social networks, Transactions on Intelligent Systems and Technology 7 (2016) 30:1–30:23. doi:10.1145/2814575.

[9] M. Ge, C. Delgado-Battenfeld, D. Jannach, Beyond accuracy: evaluating recommender systems by coverage and serendipity, in: RecSys, ACM, 2010, pp. 257–260. doi:10.1145/1864708.1864761.

[10] M. Kaminskas, D. Bridge, Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems, Transactions on Interactive Intelligent Systems 7 (2017) 2:1–2:42. doi:10.1145/2926720.

[11] A. B. Melchiorre, N. Rekabsaz, E. Parada-Cabaleiro, S. Brandl, O. Lesota, M. Schedl, Investigating gender fairness of recommendation algorithms in the music domain, Information Processing & Management 58 (2021) 102666. doi:10.1016/j.ipm.2021.102666.

[12] M. D. Ekstrand, M. S. Pera, The demographics of cool: Popularity and recommender performance for different groups of users, in: RecSys Posters, volume 1905 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017.

[13] B. Edizel, F. Bonchi, S. Hajian, A. Panisson, T. Tassa, Fairecsys: mitigating algorithmic bias in recommender systems, International Journal of Data Science and Analytics 9 (2020) 197–213. doi:10.1007/s41060-019-00181-5.

[14] M. Fuchs, W. Höpken, Applied Data Science in Tourism, Springer International Publishing, Cham, 2022, pp. 129–149. URL: https://doi.org/10.1007/978-3-030-88389-8_8.

[15] D. Herzog, L. W. Dietz, W. Wörndl, Tourist trip recommendations – foundations, state of the art and challenges, in: M. Augstein, E. Herder, W. Wörndl (Eds.), Personalized Human-Computer Interaction, de Gruyter Oldenbourg, Berlin, Germany, 2019, pp. 159–182.
[16] B. McKercher, Towards a classification of cultural tourists, International Journal of Tourism Research 4 (2002) 29–38. doi:10.1002/jtr.346.
[17] A. Yiannakis, H. Gibson, Roles tourists play, Annals of Tourism Research 19 (1992) 287–303. doi:/10.1016/0160-7383(92)90082-Z.
[18] J. Neidhardt, L. Seyfang, R. Schuster, H. Werthner, A picture-based approach to recommender systems, Information Technology and Tourism 15 (2015) 49–69. doi:10.1007/s40558-014-0017-5.
[19] D. Massimo, F. Ricci, Popularity, novelty and relevance in point of interest recommendation: an experimental analysis, Information Technology & Tourism 23 (2021) 473–508. doi:10.1007/s40558-021-00214-5.
[20] C. Zhang, H. Zhang, J. Wang, Personalized restaurant recommendation method combining group correlations and customer preferences, Information Sciences 454-455 (2018) 128–143. doi:10.1016/j.ins.2018.04.061.
[21] Y. Si, F. Zhang, W. Liu, An adaptive point-of-interest recommendation method for location-based social networks based on user activity and spatial features, Knowledge-Based Systems 163 (2019) 267–282. doi:10.1016/j.knosys.2018.08.031.
[22] L. W. Dietz, A. Sen, R. Roy, W. Wörndl, Mining trips from location-based social networks for clustering travelers and destinations, Information Technology & Tourism 22 (2020) 131–166. doi:10.1007/s40558-020-00170-6.
[23] P. Sánchez, A. Bellogín, On the effects of aggregation strategies for different groups of users in venue recommendation, Information Processing & Management 58 (2021) 102609. doi:10.1016/j.ipm.2021.102609.
[24] A. Kariryaa, I. L. Johnson, J. Schöning, B. J. Hecht, Defining and predicting the localness of volunteered geographic information using ground truth data, in: CHI, ACM, 2018, p. 265. doi:10.1145/3173574.3173839.
[25] X. Ning, C. Desrosiers, G. Karypis, A comprehensive survey of neighborhood-based recommendation methods, in: Recommender Systems Handbook, Springer, 2015, pp. 37–76. doi:10.1007/978-1-4899-7637-6_2.
[26] F. Aiolli, Efficient top-n recommendation for very large scale binary rated datasets, in: RecSys, ACM, 2013, pp. 273–280. doi:10.1145/2507157.2507189.
[27] Y. Hu, Y. Koren, C. Volinsky, Collaborative filtering for implicit feedback datasets, in: ICDM, IEEE Computer Society, 2008, pp. 263–272. doi:10.1109/ICDM.2008.22.
[28] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, BPR: bayesian personalized ranking from implicit feedback, in: UAI, AUAI Press, 2009, pp. 452–461.
[29] Y. Liu, W. Wei, A. Sun, C. Miao, Exploiting geographical neighborhood characteristics for location recommendation, in: CIKM, ACM, 2014, pp. 739–748. doi:10.1145/2661829.2662002.
[30] F. Yuan, J. M. Jose, G. Guo, L. Chen, H. Yu, R. S. Alkhawaldeh, Joint geo-spatial preference and pairwise ranking for point-of-interest recommendation, in: ICTAI, IEEE Computer Society, 2016, pp. 46–53. doi:10.1109/ICTAI.2016.0018.
[31] C. Cheng, H. Yang, I. King, M. R. Lyu, Fused matrix factorization with geographical and

social influence in location-based social networks, in: AAAI, AAAI Press, 2012.

[32] Q. Yuan, G. Cong, Z. Ma, A. Sun, N. Magnenat-Thalmann, Time-aware point-of-interest recommendation, in: SIGIR, ACM, 2013, pp. 363–372. doi:10.1145/2484028.2484030.

[33] A. Said, A. Bellogín, Comparative recommender system evaluation: benchmarking recommendation frameworks, in: RecSys, ACM, 2014, pp. 129–136. doi:10.1145/2645710.2645746.

[34] S. Vargas, P. Castells, Rank and relevance in novelty and diversity metrics for recommender systems, in: RecSys, ACM, 2011, pp. 109–116. doi:10.1145/2043932.2043955.

[35] A. Gunawardana, G. Shani, Evaluating recommender systems, in: Recommender Systems Handbook, Springer, 2015, pp. 265–308. doi:10.1007/978-1-4899-7637-6_8.

[36] H. Abdollahpouri, R. Burke, B. Mobasher, Managing popularity bias in recommender systems with personalized re-ranking, in: FLAIRS Conference, AAAI Press, 2019, pp. 413–418.

[37] L. Boratto, G. Fenu, M. Marras, The effect of algorithmic bias on recommender systems for massive open online courses, in: ECIR (1), volume 11437 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 457–472. doi:10.1007/978-3-030-15712-8_30.

[38] L. Ravi, S. Vairavasundaram, A collaborative location based travel recommendation system through enhanced rating prediction for the group of users, Computational Intelligence and Neuroscience 2016 (2016) 1291358:1–1291358:28. doi:10.1155/2016/1291358.

[39] L. W. Dietz, R. Roy, W. Wörndl, Characterisation of traveller types using check-in data from location-based social networks, in: ENTER, Springer, 2019, pp. 15–26. doi:10.1007/978-3-030-05940-8_2.