



On the effects of aggregation strategies for different groups of users in venue recommendation

Pablo Sánchez ^{*}, Alejandro Bellogín

Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain

ARTICLE INFO

Keywords:

Venue recommendation
Data augmentation
Temporal evaluation
Tourism
User types
Fairness

ABSTRACT

Suggesting new venues to be visited by a user in a specific city remains an interesting but challenging problem, partly because of the inherent high sparsity of the data available in location-based social networks (LSBNs). At the same time, in traditional recommender systems, in order to improve their performance in these sparse situations, different techniques have been proposed mainly by augmenting and aggregating the data available in different domains. In this paper, we address the problem of venue recommendation from a novel perspective: we propose two strategies to select a set of candidate cities in order to use their information when performing recommendations for the users in a specific (target) city. In this context, we categorize users into two different groups (tourists and locals) according to their movement patterns and analyze the potential biases in the recommendations received by each of these groups. We provide an experimental comparison of several recommendation algorithms in a temporal split, where we analyze two strategies to select cities and augment the available data: based on the number of interactions and based on the distance with respect to the target city. Our results show that, in general, extending the available data by proximity increases the performance of the majority of the tested recommenders in terms of relevance and coverage, with almost no change in novelty and diversity. We have found that those users belonging to the tourist group tend to obtain better results in terms of relevance. Furthermore, in general, tourists consistently exhibit different performance by some families of recommenders for other evaluation dimensions, evidencing a popularity bias in user behavior and raising potential fairness issues regarding the quality of the received recommendations. We investigate these aspects and provide methods to better understand the problem. We expect these results could provide readers with an overall picture of what can be achieved in a real-world environment.

1. Introduction

The great development of location-based social networks (LBSNs) in recent years has encouraged the research on the problem of Point-of-Interest (POI) or venue recommendation, i.e., suggesting new places to visit by analyzing the users' tastes, needs, and movement patterns (Lian et al., 2018). Foursquare and Gowalla are examples of this kind of social networks, where users record the check-ins they make to certain venues (restaurants, cinemas, hotels, etc.) and share their experiences with other users in the system (Ye, Yin, Lee, & Lee, 2011; Zhang & Chow, 2013). This information, if processed and exploited correctly, can then be used to suggest to the users new venues to visit when using a recommendation engine.

^{*} Corresponding author.

E-mail address: pablo.sanchezp@uam.es (P. Sánchez).

<https://doi.org/10.1016/j.ipm.2021.102609>

Received 16 November 2020; Received in revised form 8 March 2021; Accepted 13 April 2021

Available online 18 May 2021

0306-4573/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Since research on Recommender Systems (RS) has increased in many directions in the last years, it is important to consider some specific details of POI recommendation that differ from the traditional recommendation problem (Li, Cong, Li, Pham, & Krishnaswamy, 2015; Liu, Pham, Cong, & Yuan, 2017; Wang, Terrovitis, & Mamoulis, 2013). These include, but are not limited, to:

- **Implicit and repeated interactions:** check-in data is one type of implicit feedback, where only positive values indicating that a user has visited a venue are recorded. Nonetheless, since users may check-in at the same place several times, researchers often build frequency matrices to model these repetitions. In fact, the presence of repeated interactions has a strong effect on performance when exploited properly (Sánchez & Bellogín, 2018). Other sequence-aware recommendation areas such as music or e-commerce also exploit these repetitions for their recommendations (Quadrana, Cremonesi, & Jannach, 2018; Schedl, Zamani, Chen, Deldjoo, & Elahi, 2018); in contrast, in traditional domains such as books or movies it is normally assumed that users rate each item once (Ning, Desrosiers, & Karypis, 2015).
- **External influences:** venue recommendation is highly affected by social (user friends), temporal, and geographical influences. The latter is possibly the most important effect to consider in POI recommendation to improve the recommendation performance, as it is usually assumed that users prefer to visit venues that are close to each other (as the first law of geography states “*Everything is related to everything else, but near things are more related than distant things*” (Miller, 2004)). That is the main reason why researchers have proposed algorithms modeling explicitly the locations of venues (Lian et al., 2014; Liu, Wei, Sun, & Miao, 2014; Ye et al., 2011).
- **Sparsity:** in the RS domain, the user–item matrix is usually very sparse. However, in venue recommendation this effect is even more severe. For example, the densities of the Movielens20M and Netflix datasets are 0.539% and 1.177% respectively, while the density of the Foursquare dataset that we are using in our work is 0.0034%. It should be emphasized, however, that these density values might be, in some cases, misleading. By considering the law of geography previously stated, people living in a city would check-in along a somewhat limited subset of the venues, i.e., those of their city. Nonetheless, the sparsity problem is still key in these systems, since the number of *active* users in a city is also very small. Hence, the number of interactions that these venues receive is very low, leading to very high sparsity scenarios.

At the same time, recommenders can be improved by extending the available information with additional data. This is especially important in datasets with a high level of sparsity, where it is difficult to learn patterns of users and items due to the low number of interactions between them. Transfer (or cross-domain) learning is one of these valuable techniques that allow us to use external or additional information, mainly to improve the performance in the target domain (Lu et al., 2015; Yu, Chu, Jiang, Guo, & Gong, 2018). In the context of RS, cross-domain recommendation is a recent and active research topic, where POI recommendation has been acknowledged as a potential application domain (Zheng, 2015). However, not many experimental comparisons have been performed using cross-domain or augmentation techniques in this field with a temporal evaluation methodology. In this regard, it is important to mention that it is common to find POI recommendation proposals that perform other type of splits for evaluation, such as random partitions, cross-validation, or temporal splits per user (Liu et al., 2014; Ye et al., 2011; Yuan et al., 2016). However, we argue that these types of splits are less realistic, since in those partitions we may be predicting user interactions that occurred in the past mixed with other events that occur in the future. In addition, user tastes change over time and also there may be global trends that we would not be taking into account if a temporal partition is not used.

Moreover, besides the inherent personalized results that are expected to be received by users, these users are traditionally treated equally when measuring the performance of the recommenders. However, this is slowly changing since recently the field is paying more attention to whether users with different attributes (such as age, gender, nationality, etc.) receive the same treatment, or, in other terms, if the recommender system provides *fair* recommendations (Edizel, Bonchi, Hajian, Panisson, & Tassa, 2020; Ekstrand, Tian, Azpiazu et al., 2018). Nonetheless, these efforts are not easy to generalize to other domains, in particular because users do not share the same characteristics in different recommender systems and also because in some domains, user groups might not be easily categorized with a single attribute, even though such groups exist and evidence distinct behavior. In particular, in the tourism domain (which is typically studied with data from LBSNs), check-in data has been used to characterize four types of travelers (Dietz, Roy, & Wörndl, 2019): vacationers, explorers, voyagers, and globetrotters, thus going beyond the classical tourist roles that are usually considered (either leisure or business). Nonetheless, it is also acknowledged that most of these users are not actual tourists but local users, or, at least, that they tend to travel a limited distance from their home locations (an effect called *travel locality*) (Levandoski, Sarwat, Eldawy, & Mokbel, 2012). For this reason, in this paper we will focus on two main groups of users, locals and tourists. The first group, as said before, tends to be numerous, and sometimes this type of users are considered as “local experts”, and some approaches exploit their interactions to improve the performance of the recommendation models (Bao, Zheng, & Mokbel, 2012; Bao, Zheng, Wilkie, & Mokbel, 2015). Tourists, on the other hand, tend to spend less time in the target city and although they may have a different behavior than locals they may also be interested in the local culture (Choudhury et al., 2010; Zhao, Nie, Wang, & Chua, 2014). In any case, and independent of these uncertainties in the types of users that can be identified in these systems, we think it is important to study if the susceptible algorithms to be used in such systems are to some extent biased towards any particular user group, or if these groups are transparent to the algorithms.

Considering these issues, in this paper we analyze the effect of producing recommendations when using augmented information extracted from other cities, by exploiting different Multi-City Aggregation (MCA) strategies, either based on the number of interactions (popular cities) or by proximity with respect to the target city. In fact, LBSNs users can travel anywhere in the world and the check-ins they perform in one city can be useful to learn their profiles and used to make recommendations in other cities. For that reason we will use two simple strategies to obtain more data to make recommendations: with the first strategy (based on

popular cities) we will be able to further expand the information available to the algorithms, whereas with the second (based on geographic proximity, either by distance or by the country of the target city) we expect to exploit related interactions as it is easier to travel to nearby cities. Furthermore, we incorporate into our work an exploratory analysis with the aim to uncover any bias or effect that may inadvertently be present when making venue recommendations from LBSN data, on both state-of-the-art algorithms and when using our proposed MCA strategies. Therefore, in this work we address the following research questions:

- RQ1: Are state-of-the-art recommendation algorithms able to exploit augmented information through MCA strategies for venue recommendation?** We empirically compare a set of recommenders under two MCA strategies, where information from many other cities is incorporated into the recommendation algorithm: according to their popularity and according to the distance to the target city (with a special case considering the cities of the same country with respect to the target city). More specifically, we are interested in analyzing which is the best MCA strategy to augment the available data in terms of relevance. For this, we focus on nDCG as ranking metric and see which aggregation strategy has a larger impact in performance under a realistic temporal evaluation.
- RQ2: What is the impact that venue recommenders have on different groups of users?** We identify two major user groups, tourists and locals according to their check-in distribution, that is, only based on information that is available in any LBSN; in particular, our goal is to classify and analyze the differences and biases towards users that act as tourist in contrast to those that are locals in a particular city. We analyze both state-of-the-art algorithms and the proposed MCA strategies, in the context of venue recommendation. Thus, we explore systematic biases revealed by ranking accuracy from these approaches and hypothesize their primary reasons.
- RQ3: How do MCA strategies affect other evaluation dimensions? What is the impact of these dimensions on different groups of users?** We analyze the novelty, diversity, and coverage of the recommenders after MCA strategies are used to augment the data. We then compare these metrics against their values when no aggregation strategy is used to see how these dimensions are affected when using additional data on the recommenders. At the same time, we explore the extent to which the biases revealed in the previous research question, that are related to accuracy, are also present when measuring performance with beyond-accuracy evaluation metrics.

Our work provides a thorough comparison of two Multi-City Aggregation strategies for venue recommendation under a realistic time-aware evaluation methodology. We report the results obtained by the MCA strategies in terms of ranking accuracy, novelty, diversity, and coverage using a dataset with more than 30M interactions. Moreover, we complement these results with an extensive analysis of the effect of these strategies (together with the base performance of state-of-the-art approaches) in different types of users.

Our results indicate that proper data augmentation strategies improve the performance of some recommendation approaches, although for other recommendation algorithms, such as those based entirely on geographical information, it is more appropriate to not augment the input data since those techniques might get affected negatively by the geographical biases introduced by the MCA strategies. Furthermore, data augmentation strategies obtain such improvements at the expense of almost no change in novelty and diversity, and even improving their coverage in some cases. At the same time, our results show that tourist users systematically obtain better recommendations (at least, in terms of accuracy), which evidences a strong bias towards this type of users. We further conclude that this behavior is due to the preferences of these users being more prominent towards popular POIs in the cities, which are easier to satisfy. This result is, to the best of our knowledge, novel in the area and opens up several possibilities, as we shall show later in this work.

2. Background

The aim of Recommender Systems is to assist users through large databases and catalogs, by filtering and suggesting relevant items taking into account the users' preferences (i.e., tastes, interests, or priorities). In the RS community, a large number of algorithms have been proposed in order to learn and model the user interests. The most well-known approaches are collaborative filtering (CF) and content-based (CB) algorithms (which are usually combined creating hybrid RS) and, recently, context-aware recommenders have also emerged due to their ability to model additional information such as time, sequentiality, user location, or even the weather in order to make recommendations. However, given their pervasiveness in the literature, we will skip their definitions here and refer the reader to [de Gemmis et al. \(2015\)](#) to learn about CB systems, to [Koren and Bell \(2015\)](#), [Ning et al. \(2015\)](#) to know more about CF techniques (such as nearest neighbors and matrix factorization approaches) and, finally, to [Villegas, Sánchez, Díaz-Cely, and Tamura \(2018\)](#) for a more in-depth survey of context-aware recommenders.

The recommendation of Point-Of-Interest (POI) or venues is similar to the traditional recommendation problem, since its main objective is to recommend venues (which may belong to any category such as restaurants, museums, parks, etc.) that the user has not seen beforehand, but it exhibits some distinctive features. For example, in traditional recommendation we normally assume that the users consume an item only once, while in POI recommendation multiple visits to the same venue can be exploited when building the recommendation model ([Liu et al., 2014](#); [Sánchez & Bellogín, 2018](#)). Besides, this type of recommendation is highly influenced by spatial, social, and temporal factors ([Chiang & Huang, 2015](#); [Liu et al., 2017](#)), so most of the algorithms try to model them in order to perform recommendations.

Table 1

Jaccard coefficient between users in the training splits of each city and the corresponding MCA strategy.

Multi-city aggregation strategy (MCA)	Cities								Average
	Istanbul	Jakarta	Kuala Lumpur	Mexico City	Moscow	Santiago	São Paulo	Tokyo	
N-MCA	89.54%	67.14%	83.44%	89.36%	94.48%	85.28%	64.87%	79.13%	81.66%
C-MCA	67.64%	57.54%	66.14%	67.80%	61.35%	83.68%	34.43%	74.52%	64.14%
P-MCA	29.09%	13.63%	13.53%	9.30%	8.71%	7.73%	8.99%	10.79%	12.72%

Matrix Factorization (MF) techniques are one of the most extended approaches in the RS area because they tend to outperform other techniques such as neighbor-based algorithms (Koren, Bell, & Volinsky, 2009), moreover, they can also be easily extended with baseline predictors and temporal information (Koren & Bell, 2015). In venue recommendation these techniques are also very popular and they form the core of many approaches, such as LRT (Gao, Tang, Hu, & Liu, 2013), iRenMF (Liu et al., 2014), GeoMF (Lian et al., 2014) and Rank-GeoFM (Li et al., 2015).

Other methods besides those based on MF have also been applied in venue recommendation. For example, the USG model from Ye et al. (2011) combines three different components: a probabilistic model based on the history of visited POIs by the user to consider geographical information, CF similarities based on other users in the system (classic CF), and CF similarities based on the friends of the target user (social influence). Another interesting approach is the LORE algorithm from Zhang, Chow, and Li (2014), where Markov Chains are used to model the sequential patterns between POIs, together with social and geographical influence.

As we observe, there is a great number of venue recommendation algorithms but the basis for most of them are proposals already explored in traditional recommendation, complementing them with specific features of LBSNs (geographical, social, and temporal influence, among others). Because of this, in this paper, we analyze the impact of Multi-City Aggregation strategies in the context of venue recommendation and analyze the effect these strategies have on tourist and local users in different cities using data from a well-known LBSN, Foursquare. In the next section we define these strategies more in detail and we develop some motivation for them to work.

3. POI recommendation by MCA

To avoid the inherent problems and limitations prevalent in POI recommendation, we introduce now the concept of Multi-City Aggregation (MCA) strategies, that allow us to augment the available data used by the recommendation algorithms to suggest interesting POIs to users. The basic idea behind our proposed strategies is that, to improve recommendations over a target city, the data from multiple cities will be aggregated and exploited when training the recommendation algorithms. Since an infinite number of potential strategies may exist, we shall focus on those that maximize the overlap information between users.

More specifically, the main goal we aim to achieve, hence, if we focus on CF algorithms, would be to find highly active users in the aggregated cities so the recommendations are produced based on more information. Therefore, we consider the following possibilities:

- Geographical Nearest MCA (N): we use the n closest cities to the target city as the aggregated information. We aim to capture *cultural diversity* through their patterns,¹ while, at the same time, we keep control of the number of cities we consider. Additionally, as a special case of Geographical Nearest MCA strategy, we consider a country-based MCA strategy (C), where the aggregated information is built by all the cities of the same country as the target city. In this strategy, we assume that users tend to visit more often those cities of the same country, or, at least, that there are some kind of patterns shared among them, mostly due to a similar culture and language (Yang et al., 2016), even though some situations (large countries) may add too much noise into the model. The idea for this strategy comes from previous works in POI recommendation that use datasets with interactions belonging to a specific country or a state/province of a country, as in Li et al. (2015), Xie et al. (2016). Note that for the N-MCA there could be cities from a different country with respect to the target city while in C-MCA all selected cities will belong to the same country, even though some of them may not be so close to the target city.
- Most-popular MCA (P): we use the n cities with more check-ins as the aggregated data. This strategy allows us to test whether considering those cities that the system has more information about can be useful for the model. We hypothesize that having more information should be helpful for the recommendation algorithms, however, this strategy might also be more sensitive to noise and may not improve the user overlap (even though a high user overlap may not be sufficient to obtain a performance improvement). This strategy is also influenced by other works in POI recommendation that use datasets with check-ins from all over the world, including (Gao et al., 2013; Zhang & Chow, 2015a, 2015b).

As a first validation that these strategies might be actually helpful for POI recommendation, we show in Table 1 the percentage of common users (measured using the Jaccard coefficient) between the training set of each target city and the corresponding training set of the multiple cities being aggregated according to the different strategies described before. Using the same dataset that will be

¹ In particular, according to Yang, Zhang, and Qu (2016), the most important pattern corresponds to crowd mobility (characterized as daily activity patterns and intercity mobility), combined with the distribution of venue categories.

Table 2

Description of the temporal partition evaluated (the complete dataset, the training set, Tr, and the test set, Te), where U , I , and C denote the number of users, items, and check-ins (either with repetitions, C_R , or without, C_r). The column $\delta(C)$ denotes the density of dataset as computed by $C/(U \cdot I)$.

Check-in period	U	I	C_R	C_r	$\delta(C_R)$	$\delta(C_r)$	C_r/U	C_r/I
Apr'12–Sep'13	267k	3.7M	33.3M	15.1M	0.0034%	0.0015%	123.60	9.16
Tr: May–Oct '12	202k	1.1M	9.9M	4.8M	0.0044%	0.0021%	23.67	4.31
Te: Nov '12	150k	352k	831k	831k	0.0017%	0.0017%	5.52	2.36

used and explained later in the experiments, we compute these percentages using the Jaccard coefficient as follows (in agreement with Yang et al. (2016)):

$$\text{Common Users}(C_1, C_2) = \frac{|\mathcal{U}(C_1) \cap \mathcal{U}(C_2)|}{|\mathcal{U}(C_1) \cup \mathcal{U}(C_2)|} \quad (1)$$

where $\mathcal{U}(C)$ denotes the set of users in city C .

We observe that the N-MCA and C-MCA strategies are really useful to find more users in common, however, it is not clear the actual effect this result may have on the performance of venue recommendation algorithms, especially on those not based on nearest neighbors. At the same time, even though the P-MCA strategy does not discover many users in common, since it includes much more data to train the recommenders, it may benefit some recommendation approaches. Hence, in the experiments we shall test which approach is actually more helpful to obtain better recommendations.

In summary, and drawing related concepts available in the literature, our proposal would be similar to some techniques from cross-domain (or transfer learning) recommendation, if we consider each city as a different domain. In that way, when we augment the available information from the different cities (according to the proposed MCA strategies), we would combine data from different domains, hence performing a *cross-domain recommendation*. More specifically, according to the taxonomy presented in Cantador, Fernández-Tobías, Berkovsky, and Cremonesi (2015), it would fit in the category of merging user preferences by aggregating knowledge, since we combine multiple sources of personal preferences (basically, the check-ins from various cities and the target city). Additionally, our scenario is special and more difficult according to the literature since no item overlap exists between the domains, this is because an item (venue) will never appear in a different domain since they are unique (even two places of the same food or clothing chain will be considered different).

The main advantage of applying some kind of cross-domain in venue recommendation is that we can expand the knowledge of the recommenders with a larger number of users and items, in order to establish more relationships between them and find better patterns. However, it is not obvious how we should select such knowledge: on the one hand, we might add noise to the model and, on the other hand, such information may not be useful at all. This is exactly what we propose to study and analyze in this paper: MCA strategies that select data according to different hypotheses and criteria.

4. Experimental settings

4.1. Dataset

The experiments have been performed using the global-scale check-in dataset of Foursquare² made public by the authors of Yang et al. (2016), by capturing those check-ins posted by Twitter users. Considering the worldwide nature of this dataset, it seems impractical for us to select a training/test partition that would remain comparable for any city in the world. Therefore, we aimed to maximize the amount of data included for training the models and to test their performance, while, at the same time, the temporal patterns in each split should not belong to seasons that are too different to each other. Based on this, and starting from the original 33M check-in over 415 different cities, we created a temporal split containing 6 months of data in its training step (from May to Oct '12) and one month for test (Nov '12). It should be noted that this dataset includes POIs of any type, from restaurants, shops, or museums to landmarks. Table 2 shows more statistics of both the original dataset and the training/test split used in this paper. Additionally, as a pre-processing step, we performed a 2-core before splitting the data into training and test, which means that every user and POI has at least two interactions.

4.2. Compared baselines

We report results obtained by the following state-of-the-art recommenders grouped in different families according to the common mechanisms used to make the recommendations.

- Classic non-personalized (NP). Traditional recommendation algorithms that do not learn a profile for each user:
 - Popularity (Pop): recommender that suggests the most popular items, i.e., venues visited by more unique users.
 - Random (Rnd): random recommender.

² <https://sites.google.com/site/yangdingqi/home/foursquare-dataset>.

- Only geographical (Geo). Basic algorithms used to model only the geographical component:
 - Average Distance (AvgDis): baseline recommender that suggests the closest POIs to the user's average location. The average is computed by calculating the midpoint of the coordinates of the POIs visited by each user.
 - Kernel Density Estimation (KDE): the geographical influence component from Zhang et al. (2014). It models a probability distribution over a two-dimensional space using Kernel Density Estimation for every user.
- Classic collaborative filtering (CF-NN). Traditional recommendation algorithms based on nearest neighbors:
 - Item neighborhood (IB): a k -NN recommender with an item-based approach (Ning et al., 2015). We use a not-normalized version since it shows better ranking performance (Aioli, 2013).
 - User neighborhood (UB): a k -NN recommender with a user-based approach (again, without normalization) (Aioli, 2013; Ning et al., 2015).
- Classic matrix factorization (CF-MF). Traditional recommendation algorithms based on matrix factorization approaches:
 - Bayesian Personalized Ranking (BPR): the Bayesian Personalized Ranking from Rendle, Freudenthaler, Gantner, and Schmidt-Thieme (2009) using a matrix factorization technique as provided in the MyMediaLite library.³
 - Alternate Least Squares (ALS): a matrix factorization (MF) approach as described in Hu, Koren, and Volinsky (2008) that uses Alternate Least Squares in the minimization formula.
- POI models (POI). State-of-the-art POI recommendation algorithms:
 - Geographical Bayesian Personalized Ranking (GeoBPR): a POI recommendation approach that uses BPR to optimize the model (Yuan et al., 2016). It incorporates the geographical influence by assuming that users prefer to visit close rather than remote POIs with respect to the ones that the user has already visited.
 - Instance-Region Neighborhood Matrix Factorization (IRenMF): weighted MF method proposed in Liu et al. (2014). We selected this approach because, according to the comparison presented in Liu et al. (2017), IRenMF was very competitive with a lower execution time with respect to other models, such as GeoMF (Lian et al., 2014), or LFBCA (Wang et al., 2013), which agrees with some preliminary experiments we performed in our dataset.
 - Ranking-based Geographical Factorization Method (Rank-GeoFM): a ranking-based matrix factorization approach model proposed in Li et al. (2015) that uses an additional latent matrix for the users to model the geographical influence by exploiting the neighboring POIs (by geographical distance) with respect to the target POI.
- Hybrid POI recommendation models (H-POI). POI recommendation algorithms whose final score is the combination of two or more independent algorithms:
 - Probabilistic Matrix Factorization with Multi-center Gaussian Model (FMFMGM): is a fusion proposed in Cheng, Yang, King, and Lyu (2012) that combines the Multi-center Gaussian Model technique (MGM) with Probabilistic Matrix Factorization (PMF).
 - Geographical, Social, and Categorical correlations (GeoCFCa): a hybrid POI recommendation model based on the technique proposed in Zhang and Chow (2015a) that combines the geographical influence using a two-dimensional KDE and the social and categorical influences modeled by two different power-law distributions. As in the Foursquare dataset the social information is not available for all the users, we decided to use a k -NN algorithm as a substitute for the social component; because of this, we changed its name from the original GeoSoCa to GeoCFCa.
 - Popularity, Geographical, and user-based Neighborhood (PGN): a hybrid approach similar to the USG model proposed in Ye et al. (2011) that combines the Pop, UB, and AvgDis recommenders described before. It basically aggregates the scores of every item provided by each of the recommenders, after normalizing its values by the maximum score of each method.

In order to make a fair comparison among all the evaluated baselines, we removed repetitions in a user basis for some classic algorithms. When repetitions are allowed, we either aggregated them as frequencies or keep all the repeated interactions. This means that we kept three versions of the training set: with and without check-in frequencies but where each user-item pair only appeared once, and another scenario where the user-item pairs might be repeated. More specifically, the following POI recommendation algorithms could exploit the frequency of users when visiting a specific venue: AvgDis, IRenMF, FMFMGM, RankGeoFM and GeoCFCa; whereas KDE is the only recommender that uses the training set with repetitions.

For every recommender except the IRenMF and BPR algorithms, we used the RankSys library (Castells, Hurley, & Vargas, 2015); for IRenMF we used the implementation provided by Liu et al. (2014), available here⁴; for BPR, as stated before, we used the implementation from the MyMediaLite library; for the Geo family, we used our own versions of the algorithms; for Rank-GeoFM and GeoBPR, we implemented our own versions on top of RankSys, based on the implementation provided by Librec⁵ for the former,

³ <http://www.mymedialite.net>.

⁴ <http://spatialkeyword.sce.ntu.edu.sg/eval-vldb17/>.

⁵ <https://www.librec.net/>.

Table 3

Parameters used with the evaluated recommenders. SetC and SetJ stand for SetCosine and SetJaccard. Parameters optimized using Precision@5.

Recommender	Parameters
Rnd	None
Pop	None
AvgDis	None
KDE	None
BPR	Factors={10, 50, 100}, Iter=50, LearnRate=0.5, RegJ=RegU/10, BiasReg={0, 0.5, 1}, RegU=RegI={0.0025, 0.001, 0.005, 0.01, 0.1}
ALS	Factors={10, 50, 100}, $\alpha = \{0.1, 1, 10\}$, $\lambda = \{0.1, 1, 10\}$
UB	Sim={SetC, SetJ}, $k = \{5, 10, \dots, 100\}$
IB	Sim={SetC, SetJ}, $k = \{5, 10, \dots, 100\}$
GeoBPR	MaxDist={1, 4}, Factors={10, 50, 100}, BiasReg={0, 0.5, 1}, Iter=50, RegU=RegI={0.0025, 0.001, 0.005, 0.01, 0.1}, LearnRate=0.05
IRenMF	$k = 10$, Clusters=50, $\lambda_1 = \lambda_2 = 0.015$, Factors={50, 100}, $\alpha = \{0.4, 0.6\}$, $\lambda_3 = \{0.1, 1\}$
Rank-GeoFM	Factors={10, 50, 100}, $k = \{10, 50, 100, 200\}$, $\alpha = \{0.1, 0.2\}$, $c=1$, $\text{dec}=1$, $\epsilon=0.3$, boldDrv=true, Iter=120, LearnRate=0.001, mRate=0.001
FMFMGM	MGM: $\alpha = \{0.2, 0.4\}$, $\theta = \{0.02, 0.1\}$, dmax=15. PMF: Iter=30, Factors={50, 100}, $\alpha_2 = \{20, 40\}$, $\beta=0.2$, LearnRate=0.0001, sigmoid=false
GeoCFCa	Sim=SetJ, $k=100$
PGN	$k = 100$, Similarity=SetJ

Table 4

Optimal parameters of models for each city. The order of the presented parameters is: for UB and IB, similarity and neighborhood size; for BPR, factors, RegU, BiasReg; for ALS, factors, α , λ ; for IrenMF, factors, α , λ_3 ; for Rank-GeoFM, factors, neighborhood size, α ; for GeoBPR, maxDist, factors, RegU, BiasReg; for FMFMGM, α , θ , Factors, α_2 .

City	UB	IB	BPR	ALS	IRenMF	Rank-GeoFM	GeoBPR	FMFMGM
IST	SetJ, 90	SetC, 100	50, 0.001, 0	10, 10, 10	100, 0.4, 1	100, 50, 0.1	4, 100, 0.001, 0	0.4, 0.02, 100, 20
JAK	SetJ, 100	SetC, 80	100, 0.0025, 0	10, 10, 10	100, 0.4, 1	100, 200, 0.2	1, 100, 0.001, 0	0.2, 0.02, 100, 20
KUA	SetJ, 100	SetJ, 100	50, 0.01, 0	10, 10, 10	100, 0.4, 1	100, 50, 0.2	1, 50, 0.001, 0	0.4, 0.02, 100, 20
MEX	SetJ, 100	SetJ, 100	100, 0.01, 0	10, 10, 10	100, 0.4, 1	100, 100, 0.2	1, 100, 0.001, 0	0.4, 0.1, 100, 20
MOS	SetC, 100	SetJ, 100	100, 0.01, 0	50, 10, 1	100, 0.4, 1	100, 200, 0.1	1, 50, 0.0025, 1	0.4, 0.1, 100, 20
SAN	SetJ, 90	SetJ, 80	50, 0.005, 0	10, 10, 10	100, 0.4, 1	100, 100, 0.1	1, 50, 0.001, 0	0.4, 0.02, 100, 20
SAO	SetJ, 100	SetJ, 100	100, 0.1, 0	50, 10, 0.1	100, 0.6, 0.1	100, 50, 0.2	1, 100, 0.001, 1	0.4, 0.02, 100, 20
TOK	SetJ, 80	SetC, 80	50, 0.1, 0	10, 10, 10	100, 0.4, 1	100, 10, 0.1	4, 10, 0.001, 0	0.4, 0.02, 100, 20

and on the code provided by [Han and Yamana \(2020\)](#) for the latter. Finally, we also implemented our own versions of GeoCFCa and FMFMGM based on the code provided by [Liu et al. \(2017\)](#) for both of them and [Han and Yamana \(2020\)](#) for GeoCFCa.

The similarities used in the k -NN recommenders are based on set operations as the data does not include ratings: SetJ is the well-known Jaccard Index and SetC is based on the similarity proposed in [Aiolfi \(2013\)](#), always considering that the similarities between users/items are symmetrical. We want to note that the algorithms based on matrix factorization, since they start from a random initialization and the number of iterations is also a parameter, they may obtain different results each time they are executed. This would affect the following algorithms: FMFMGM, IrenMF, RankGeoFM, ALS, BPR and GeoBPR.

Source code to replicate these experiments can be found in the following Bitbucket repository: [PabloSanchezP/TempCDSeqEval](#)

4.3. Experimental setup

Based on the temporal split presented in [Table 2](#), we decided to focus on the eight largest cities in terms of number of check-ins: Istanbul (IST), Jakarta (JAK), Kuala Lumpur (KUA), Mexico City (MEX), Moscow (MOS), Santiago (SAN), São Paulo (SAO) and Tokyo (TOK). Hence, we will compare the recommendations produced using the information of the training set of each of these eight cities independently (Single City) with the recommendations obtained by augmenting the data from any other cities in the dataset according to the proposed strategies.

To evaluate the recommenders, we applied a common methodology in the area called TrainItems ([Said & Bellogín, 2014](#)), where only the venues that appear in the training set of each target city are considered as candidates, except the ones already rated by the user. Besides, we filter out in the test set all interactions that appear in the training set, so the test set is only composed by new preferences. Moreover, since the performance metrics we use are not well-defined when repetitions exist in the data, we eliminate all repeated interactions from the test set, simulating that each user will visit each POI only once. Then, we compare the recommendations generated by the different algorithms when using different training information according to the presented MCA strategies: multiple cities selected based on the geographically nearest cities (N-MCA and C-MCA) and based on most popular cities

(P-MCA), that is, the eight aforementioned cities. We want to emphasize that regardless of the data used to augment and train the algorithms, the test set and hence the candidate POIs will always belong to the target city. The N closest cities and the rest of the cities of the same country to each target city are reported in [Appendix](#).

As it is standard in recent RS literature, we use an array of metrics to test the performance of the recommenders in terms of ranking accuracy, novelty, diversity, and coverage. For accuracy, we will make special emphasis on normalized Discounted Cumulative Gain (nDCG), although results in terms of Precision and Recall will also be shown [Gunawardana and Shani \(2015\)](#). The optimal parameters (shown in [Table 4](#)) were selected according to the Precision@5 values obtained for the tested parameters presented in [Table 3](#) in the scenario when no MCA strategy (SC, from Single City) is applied. This means that we use the optimal parameters found in that case and apply the same values for the scenarios when either MCA strategy is used. For novelty, diversity, and coverage we present results for the following metrics:

- EPC (expected popularity complement) gives a higher value to those items that are less popular to account for the recommendation novelty ([Vargas & Castells, 2011](#)). This metric can be formulated as follows:

$$EPC = \frac{1}{|R_u|} \sum_{i \in R_u} \left(1 - \frac{|U_i|}{U^*}\right) \quad (2)$$

where R_u denotes the recommendation list of a user.

- Gini (or sales diversity) takes into account how distributed the recommended items are to measure the recommendation diversity ([Castells et al., 2015](#)). We use the following formulation:

$$\text{Gini} = 1 - \frac{1}{|I| - 1} \sum_{k=1}^{|I|} (2k - |I| - 1) p(i_k | s) \quad (3)$$

$$p(i | s) = \frac{|\{u \in U^* | i \in R_u\}|}{\sum_{j \in I} |\{u \in U^* | j \in R_u\}|} \quad (4)$$

- ISC (item space coverage) measures the percentage of unique items that a recommender returns ([Gunawardana & Shani, 2015](#)). This metric can be formulated as:

$$\text{ISC} = \frac{\left| \bigcup_{u \in U^*} R_u \right|}{|I_{tr}|} \quad (5)$$

where I_{tr} denotes the set of items in the training set.

- USC (user space coverage): measures the percentage of unique users that a recommender can provide recommendations (U_{rec}) with respect to the number of users in the test set. It can be computed as follows:

$$\text{USC} = \frac{|U_{rec}|}{|U_{test}|} \quad (6)$$

For all metrics, the higher the value, the higher the relevance/novelty/diversity of the evaluated algorithm. Additionally, unless stated otherwise, all metrics are reported at a cutoff of 5. Finally, in order to clarify some aspects about the evaluation performed, we describe now in more detail how we compute the evaluation metrics. In particular, we used a standard procedure in the Information Retrieval area in which the value of each metric for every recommender is normalized according to the number of users it is able to provide recommendations for. This means that if, for instance, there is a recommender A that has a user coverage of 100 users and a recommender B with a value of 120 users, each metric for recommender A and B will be normalized by 100 and 120, respectively.

5. Analysis of the results

In this section, we describe the results obtained when applying the evaluation methodology presented in the previous section. First, in [Section 5.1](#), we analyze the performance of venue recommenders under the proposed MCA strategies. Later, in [Section 5.2](#), we show the impact of the recommenders on two groups of users (tourist and locals). Finally, we present in [Section 5.3](#) performance results in other evaluation dimensions such as novelty and diversity.

5.1. Effect of MCA strategies in accuracy evaluation metrics

In this section, we aim to answer the first research question described at the beginning of the paper, that is: *Are state-of-the-art recommendation algorithms able to exploit augmented information through MCA strategies for venue recommendation?* With this goal in mind, we analyze which POI recommendation algorithms tend to improve or deteriorate their performance under the different MCA strategies proposed.

First, we present in [Table 5](#) the results of each recommender (for all families) in each of the eight cities. In this case, no MCA strategies are applied, so both the training and test sets correspond to the target city. The most noticeable result that we observe in this table is the low values obtained by all the recommenders. These low values are mostly due to the high sparsity of the data (see [Tables 2](#) and [7](#) for additional statistics about the cities and aggregation strategies used), together with the fact that we are using a temporal split, which makes the recommendation task even more difficult, since, for instance, a small subset of the few items a

Table 5

Performance results (nDCG@5) when no MCA strategy is used. In bold, we show the highest value for each city in each family and we show with a dagger the highest value in each city.

Family	Rec	IST	JAK	KUA	MEX	MOS	SAN	SAO	TOK
NP	Pop	0.054	0.066	0.066	0.041	0.027	0.051	0.053	0.069
	Rnd	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Geo	AvgDis	0.001	0.001	0.001	0.001	0.002	0.001	0.002	0.001
	KDE	0.003	0.004	0.004	0.006	0.006	0.005	0.008	0.005
CF-NN	IB	0.059	0.035	0.042	0.013	0.017	0.026	0.015	0.048
	UB	0.073	0.070	0.073	0.044	0.037	0.053	0.049	0.069
CF-MF	BPR	0.053	0.057	0.064	0.037	0.029	0.047	0.035	0.066
	ALS	0.070	0.066	0.066	† 0.047	0.039	0.050	0.048	0.059
POI	GeoBPR	0.064	0.067	† 0.073	0.046	† 0.044	0.052	0.052	0.068
	IRenMF	† 0.074	† 0.071	0.072	0.042	0.035	0.049	0.044	0.065
	RankGeoFM	0.059	0.044	0.051	0.027	0.019	0.031	0.025	0.036
H-POI	FMFMGM	0.060	0.050	0.060	0.029	0.028	0.033	0.034	0.061
	GeoCFCa	0.033	0.026	0.029	0.017	0.016	0.013	0.017	0.026
	PGN	0.067	0.067	0.070	0.043	0.032	† 0.054	† 0.057	† 0.070

user may have in her test set may not appear in the training set at all. Moreover, another aspect that makes it less likely to achieve high accuracy values is that, because of the temporal split, there could be some new users that did not appear in the training set, so in those cases it would be impossible to produce recommendations using any personalized model.

Regarding the recommenders, the AvgDis recommender is the second worst algorithm (after Rnd) and followed by KDE, which evidences that simply modeling the user by the geographical coordinates of her (most frequent) visited venues is not enough to predict her future interactions. Additionally, we observe that the IRenMF approach, even though it remains very competitive, it is not always the optimal recommender across all the cities, somewhat in contradiction with reported experiments in previous works (Liu et al., 2017, 2014). We attribute this behavior to the following reasons: first, its claimed superior performance was only tested using a random split in Liu et al. (2014) instead of a temporal evaluation as we do here; and second, classical recommendation algorithms such as Pop or standard CF approaches were neglected in Liu et al. (2017) which, together with our previous discussion, definitely disturbs such comparisons.

Similarly, Rank-GeoFM performs poorly, only slightly superior to IB. This could again be explained by the different conditions of the experiments in the original paper (Li et al., 2015) with respect to ours. For example, in the original paper the authors tested their algorithm only in one city on Foursquare (Singapore) and in two states in Gowalla (Nevada and California), while we have selected 8 different cities on Foursquare. On the other hand, GeoBPR shows very good results for some cities (such as KUA and MOS), although not as good in others (see IST). In this context, we must take into account that for all the recommenders that use matrix factorization techniques, there is a large number of configurable and tunable parameters, making it very difficult (and costly) to find the best configuration in all the situations.

For the rest of the recommenders, we observe that UB is one of the best approaches for most of the selected cities, usually very close to the optimal one. However, it should be taken into account that all personalized recommenders (except PGN) have less user coverage than the NP family and we may find users in the test set that have not rated any item in the training set. Later, in Section 5.3 we shall discuss this aspect again. Additionally, it is interesting to note the relatively high performance of PGN, since it is able to beat the rest of the baselines in many cities, despite its simplicity and the fact that we did not perform any parameter tuning. A possible explanation for this effect is the popularity bias, which is an important component of the PGN algorithm. As we observe in these results, the pure popularity recommender (Pop) obtains a very good performance, being able to surpass other more complex algorithms such as IB or BPR.

We now present in Table 6 the results for all cities and the best recommender (according to nDCG@5) of each family when the two proposed MCA strategies are used: most-popular MCA (P), geographically nearest MCA (N), and an MCA (C) based on the cities of the same country. In the first case, for every city, the training set is built by aggregating the training data from the eight most-popular cities; in the second one, the training set is built by taking the geographically nearest 7 cities with respect to the target city, so the number of cities under consideration is comparable to that of P-MCA, and finally for the country-based strategy, the training set is built using the check-ins of all the cities belonging to the same country of the target city.

In all scenarios the test set corresponds to the one of the target city. It should be noted that we do not show results for the non-personalized family as the MCA strategies do not affect this type of recommenders. The reason for this is that, even though some models may be trained with data from different cities, we only allow to recommend items from the target city; hence, the most popular items of a given city will remain the same regardless of the set of cities used as source domain. Similarly, the random recommender will always give a random score for each item of the target city notwithstanding the aggregation strategy used.

Thus, Table 6 shows the relative improvement with respect to the base performance of each algorithm when no MCA strategy is used to augment the training information – that is, the performance of that algorithm when only information from the target city is used, which corresponds to the results shown in Table 5. We observe that classical CF algorithms (CF-NN and CF-MF) are able to exploit quite successfully the augmented information using the P-MCA strategy, although the CF-MF often has a lower performance with respect to the base scenario (column SC, from Single City). Moreover, CF-NN always evidences a positive improvement under

Table 6

Performance in terms of nDCG@5 when augmented information is used for training. The improvement in performance with respect to Single City (SC) is represented as Δ (%).

City	Family	SC	MCA			Δ MCA (%)		
			N	C	P	Δ N (%)	Δ C (%)	Δ P (%)
IST	Geo	0.003	0.003	0.002	0.003	-14.16	-27.72	-0.90
	CF-NN	0.073	0.073	0.075	0.073	0.32	3.83	0.35
	CF-MF	0.070	0.071	0.073	0.068	1.98	4.79	-3.41
	POI	0.074	0.076	0.077	0.072	2.88	3.46	-3.25
	H-POI	0.067	0.068	0.071	0.068	1.61	6.17	0.91
JAK	Geo	0.004	0.003	0.003	0.004	-19.73	-24.52	1.83
	CF-NN	0.070	0.075	0.078	0.071	6.68	10.46	0.37
	CF-MF	0.066	0.070	0.072	0.060	6.31	9.34	-8.50
	POI	0.071	0.076	0.077	0.064	8.16	9.39	-8.63
	H-POI	0.067	0.070	0.072	0.068	4.75	6.86	0.63
KUA	Geo	0.004	0.003	0.003	0.004	-37.28	-41.48	-0.73
	CF-NN	0.073	0.076	0.078	0.073	4.13	6.20	0.29
	CF-MF	0.066	0.075	0.077	0.065	13.79	17.14	-1.55
	POI	0.073	0.075	0.075	0.062	2.20	2.58	-15.48
	H-POI	0.070	0.072	0.073	0.070	2.12	3.65	0.02
MEX	Geo	0.006	0.005	0.004	0.005	-13.79	-26.19	-1.37
	CF-NN	0.044	0.045	0.047	0.045	1.62	7.17	1.24
	CF-MF	0.047	0.045	0.045	0.037	-5.03	-3.95	-22.07
	POI	0.046	0.044	0.043	0.031	-3.14	-4.64	-31.44
	H-POI	0.043	0.044	0.046	0.044	2.21	7.13	1.33
MOS	Geo	0.006	0.005	0.005	0.006	-13.91	-22.09	-0.56
	CF-NN	0.037	0.038	0.041	0.037	2.53	10.83	0.33
	CF-MF	0.039	0.039	0.041	0.036	1.84	6.11	-7.70
	POI	0.044	0.041	0.038	0.022	-7.60	-13.54	-49.03
	H-POI	0.032	0.033	0.036	0.032	0.78	10.87	0.06
SAN	Geo	0.005	0.003	0.004	0.005	-36.08	-34.94	-1.40
	CF-NN	0.053	0.060	0.064	0.054	12.98	19.34	0.86
	CF-MF	0.050	0.060	0.062	0.046	20.21	24.22	-7.87
	POI	0.052	0.057	0.055	0.038	9.46	6.74	-27.19
	H-POI	0.054	0.059	0.060	0.055	8.68	10.16	1.14
SAO	Geo	0.008	0.007	0.006	0.008	-12.00	-19.03	-0.49
	CF-NN	0.049	0.056	0.060	0.049	15.42	23.91	-0.22
	CF-MF	0.048	0.056	0.058	0.047	15.23	19.78	-2.09
	POI	0.052	0.047	0.040	0.031	-10.26	-22.55	-40.43
	H-POI	0.057	0.057	0.058	0.057	0.41	2.06	0.53
TOK	Geo	0.005	0.003	0.003	0.004	-42.38	-45.74	-3.43
	CF-NN	0.069	0.073	0.074	0.069	5.38	7.41	-0.20
	CF-MF	0.066	0.066	0.065	0.062	0.93	-0.63	-6.13
	POI	0.068	0.071	0.070	0.064	3.90	2.89	-6.30
	H-POI	0.070	0.073	0.074	0.070	4.89	6.15	-0.16

the N-MCA and C-MCA strategies. We argue this trend for the CF techniques is related to whether the new users – it should be noted that the new items found in the augmented training set will never have overlap with the target items, since when using information from other cities the POIs will always be different – have some level of interaction with the target city. In this sense, when combining information from nearby cities it is more likely to find similar users with useful suggestions or learning relevant latent representations more related to the target items. Additionally, having more data available does not guarantee better recommendations, since the MF approaches from CF-MF and POI families tend to deteriorate their performance under the P-MCA strategy.

On the other hand, the performance of the Geo family is always worse when using any of the MCA strategies. The reason for this might be quite obvious, since considering other cities to compute a new user's centroid will certainly move such centroid far away from the target city, which is not useful when we are only interested in recommending venues inside of that specific city. Nevertheless, it is interesting to observe that the POI family, which includes a geographical component, also benefits from the MCA strategies in some scenarios. For example, in IST, JAK, KUA, SAN, and TOK using both N-MCA and C-MCA strategies, these algorithms obtain a better performance than in the Single City scenario. However, when using the P-MCA strategy, the performance of this family is always worse. This result is particularly interesting since in some works researchers perform experiments using datasets with information from several cities grouped together (Gao, Tang, & Liu, 2015; Manotumruksa, Macdonald, & Ounis, 2019; Zhang & Chow, 2013), and as we can observe, this can affect negatively the performance of some models.

From the perspective of the MCA strategies, we observe that the performance improvements obtained when using the P-MCA strategy is usually negligible. In general, most of the improvements when using this strategy are very close to zero and, for many of the city-recommender family combinations, extremely negative. At the same time, N-MCA and C-MCA usually produce larger

Table 7

Statistics for training splits of the reported cities and MCA strategies used. Notation as in Table 2. The last two columns show the amount of information in the MCA strategies that was already included in the target city C .

City		U	I	C_R	C_r	$\frac{C_R}{ U I }$	$\frac{C_r}{ U I }$	$\frac{C_R(C)}{C_R}$	$\frac{C_r(C)}{C_r}$
IST	SC	23k	40k	668k	392k	0.072%	0.042%	100.0%	100.0%
	N-MCA	26k	51k	784k	458k	0.060%	0.035%	85.21%	85.70%
	C-MCA	34k	88k	1.2M	691k	0.039%	0.023%	56.62%	56.79%
JAK	SC	11k	39k	347k	182k	0.082%	0.043%	100.0%	100.0%
	N-MCA	16k	80k	678k	354k	0.052%	0.027%	51.13%	51.52%
	C-MCA	19k	104k	861k	441k	0.044%	0.023%	40.31%	41.40%
KUA	SC	11k	28k	312k	170k	0.102%	0.056%	100.0%	100.0%
	N-MCA	13k	63k	642k	341k	0.078%	0.042%	48.62%	49.93%
	C-MCA	16k	87k	856k	438k	0.061%	0.031%	36.43%	38.79%
MEX	SC	7k	27k	285k	143k	0.144%	0.073%	100.0%	100.0%
	N-MCA	8k	35k	344k	172k	0.119%	0.059%	82.80%	83.43%
	C-MCA	11k	55k	506k	248k	0.084%	0.041%	56.34%	57.91%
MOS	SC	7k	29k	304k	153k	0.150%	0.075%	100.0%	100.0%
	N-MCA	7k	33k	328k	164k	0.137%	0.068%	92.58%	93.38%
	C-MCA	11k	64k	584k	279k	0.081%	0.039%	52.08%	54.64%
SAN	SC	6k	25k	324k	130k	0.211%	0.085%	100.0%	100.0%
	N-MCA	7k	36k	433k	173k	0.168%	0.067%	74.92%	75.39%
	C-MCA	7k	38k	455k	182k	0.162%	0.065%	71.23%	71.72%
SAO	SC	7k	28k	294k	120k	0.145%	0.059%	100.0%	100.0%
	N-MCA	11k	50k	491k	195k	0.089%	0.035%	59.76%	61.66%
	C-MCA	21k	117k	1.1M	446k	0.047%	0.018%	25.65%	26.90%
TOK	SC	9k	29k	328k	164k	0.133%	0.067%	100.0%	100.0%
	N-MCA	11k	60k	631k	301k	0.097%	0.046%	51.97%	54.40%
	C-MCA	11k	70k	705k	337k	0.088%	0.042%	46.48%	48.47%
All	P-MCA	80k	245k	2.9M	1.5M	0.015%	0.007%	–	–

improvements with less training data involved, since those cities that belong to the same country or are geographically nearest to the target city always include less check-ins than the originally selected cities, which were the most popular ones in our dataset (see Table 7 for more details). Even if these results seem to confirm that better data is more useful than more data, we now analyze these effects in more detail.

To properly understand which of the MCA strategies are more suitable to augment the data available for recommendation, we include in Table 7 the sparsity of each resulting augmented training set, together with the amount of information already included in the target city with respect to each aggregation (last two columns). Based on these statistics, we infer that the user overlap of each MCA strategy (reported in Table 1) is not the only factor to consider in the success of the proposed data augmentation approaches. For instance, the three cities with more user overlap (IST, MEX, and MOS, with more than 82%) also show high ratios of $C_{r(C)}/C_r$ when N-MCA is used, which means that such aggregation strategy incorporates very little additional information with respect to the original training data, resulting in a more sparse dataset (to be expected from any MCA strategy, due to the large amount of new items and users added) but where most of the interactions come from the same city. Going back to the results in Table 6, we observe that the N-MCA strategy is less useful (for the CF-MF approaches in particular, but also for the classical CF-NN methods) essentially when such ratio is too large, since this means that the original and augmented training splits are very similar and, hence, the performance improvement would be minimal.

Thus, we are able to answer our first research question: we have seen that some classic recommenders (i.e., nearest neighbors and matrix factorization) are able to benefit from the augmented information if the cities used to define the Multi-City Aggregation strategy are selected properly. In particular, we conclude that selecting cities by proximity (and, as a special case, by country) has a greater benefit than selecting them by the amount of information they contain (popularity). However, other approaches more tailored for venue recommendation may decrease their performance when exploiting knowledge from other domains. This is especially noticeable in strategies that give great importance to geographical influence whenever the number of common users is negligible, as in the P-MCA proposed method. This opens up the possibility of using alternative cross-domain techniques that may benefit other algorithms that are not so dependent on user overlap, such as item similarity models like SLIM, FISM, or those based on embeddings (Kabbur, Ning, & Karypis, 2013; Ning & Karypis, 2011). However, we leave as future work the analysis of these similarity models for POI recommendation together with alternative data augmentation techniques.

5.2. Analyzing venue recommender systems on different groups of users

As we have already mentioned previously, in the tourism domain it is possible to characterize different types of users. In particular, we define two groups following the work (Choudhury et al., 2010): tourists and locals. More specifically, we have established that those users whose check-ins exist in the same city for more than 21 days are considered locals, and the rest are

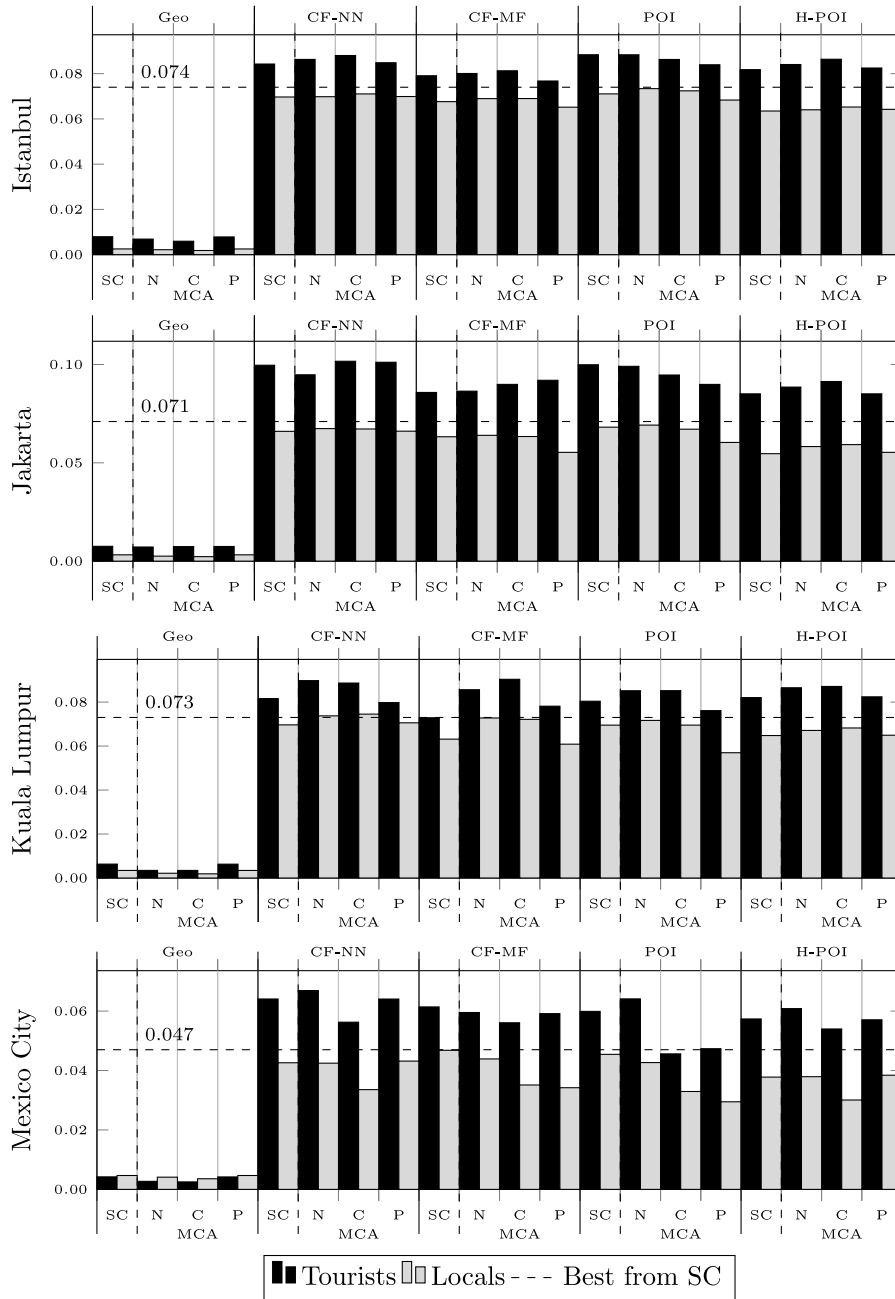


Fig. 1. Results of tourists and local users in Istanbul, Jakarta, Kuala Lumpur, and Mexico City in terms of nDCG@5. Dashed line indicates the performance of the best recommender in every city, as shown in Table 5. Labels SC, N, C, and P in the x-axis represent the single-city (baseline) configuration, N-MCA, C-MCA, and P-MCA strategies respectively.

considered tourists. However, to avoid noisy or non-human behavior, we filtered out in the test set for both groups those users who have performed three or more consecutive check-ins with a temporal difference smaller than 60 s, since they can be considered bots as in previous works (Palumbo, Rizzo, Troncy, & Baralis, 2017). These so-called bots do not count either as tourists or locals (hence, in this section they are completely ignored), even though their performance is considered whenever the global performance is measured (that is, in those tables or figures that appear in the rest of the paper). However, these unusual consecutive check-ins are not always caused by bots. It may also be due to bugs in the application when recording interactions. Therefore, we will classify

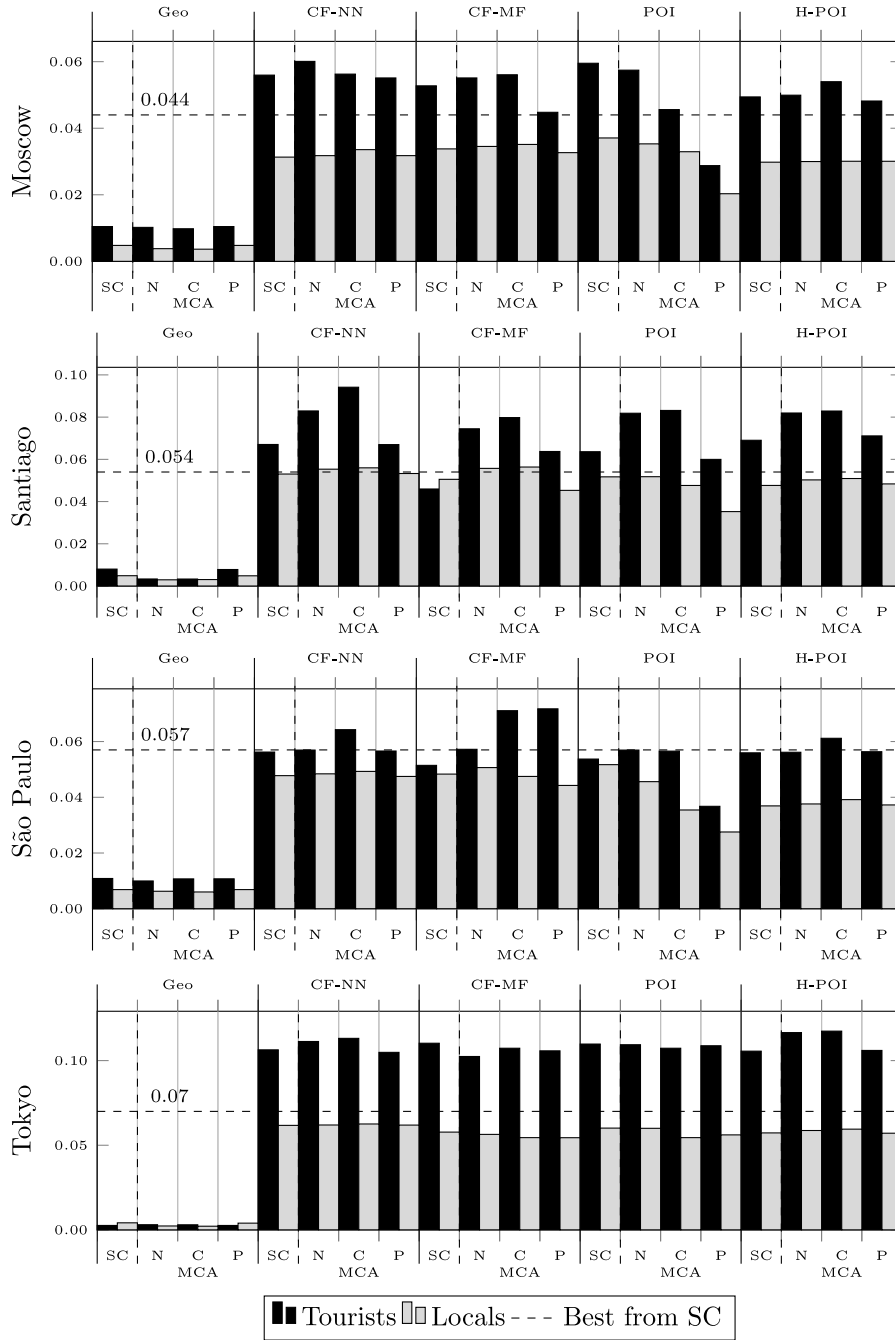


Fig. 2. Results of tourists and local users in Moscow, Santiago, São Paulo, and Tokyo in terms of nDCG@5. Rest of notation as in Fig. 1.

these users as outliers. Besides, as the data of these users may be useful for the rest of the recommenders, we keep the interactions of these users in the training set.

Based on this, in Figs. 1 and 2 we contrast the results of each type of users in terms of nDCG@5 for all cities when no MCA strategy is used (SC) and when the MCA strategies presented in the previous experiment are used (N, for N-MCA, C, for C-MCA and P, for P-MCA). We observe that for all cases (except in the Geo family in Mexico City and Tokyo), tourists obtain significantly better results than locals. We hypothesize this may be attributed to tourists having a more similar behavior in common among the users in the same group: for example, when someone visits Paris, regardless of where they come from, they are more likely to visit touristic

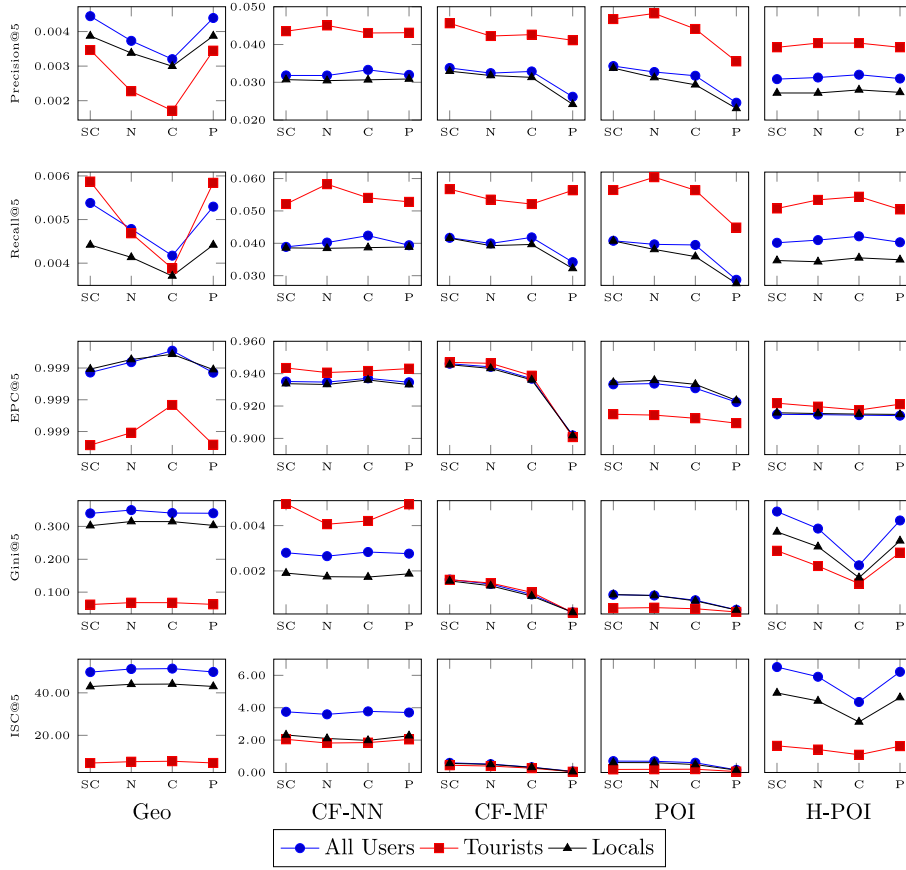


Fig. 3. Results of tourists (10.42% of the users) and local (71.60% of the users) users in Mexico City in terms of accuracy metrics (Precision, Recall), novelty (EPC) and diversity (Gini, ISC). Labels in the x-axis as in Fig. 1.

venues such as the Eiffel Tower or the Louvre museum rather than some suburban neighborhoods in the city. On the other hand, locals are probably more heterogeneous, and hence, more different behaviors are aggregated in the same group, making it much more difficult to the recommendation algorithms to guess their preferences correctly. Besides, the number of tourist users, because of its definition, tends to be smaller than local users, which helps to obtain more coherent user groups.

Consistent with the results discussed in the previous section, the MCA strategies by proximity (N-MCA and C-MCA) obtains better performance and, in general, improves the base results more than the strategy based on popularity for both types of users in most cities. There are some exceptions, as in the case of São Paulo for tourists, where the performance improvement of the P-MCA strategy is striking. However, the general trend is that this strategy is outperformed by both N-MCA and C-MCA; we note even some cases where they produce worse performance than the base scenario, for example in Moscow for most recommendation families or in Kuala Lumpur for the CF-NN and POI families.

Since we observe no different behavior with respect to the user groups between using MCA strategies or not, we come to the conclusion that venue recommenders evidence a strong bias towards tourist users, in particular, this group of users seem to be much easier to recommend. As an answer to RQ2 (*What is the impact that venue recommenders have on different groups of users?*), we summarize that every recommendation family except the basic geographical algorithms improve their results when analyzing the subset of tourists in isolation. In agreement with the previous research question, the N-MCA and C-MCA strategies are also beneficial in this case, obtaining much better results, in general, than P-MCA.

5.3. Effect of MCA strategies on beyond-accuracy evaluation metrics

An important aspect that is sometimes ignored when evaluating recommendation algorithms is finding a good balance between novelty, diversity and accuracy (Gunawardana & Shani, 2015); this is what we analyze in this section. For this, we present in Figs. 3, 4 and 5 the results for the recommendation families used before using all the metrics presented in Section 4.3 for the cities of Mexico City, Santiago and Tokyo, that is, Precision, Recall, EPC for novelty, and Gini and ISC for diversity. We complement these results with user coverage of all cities in Table 8.

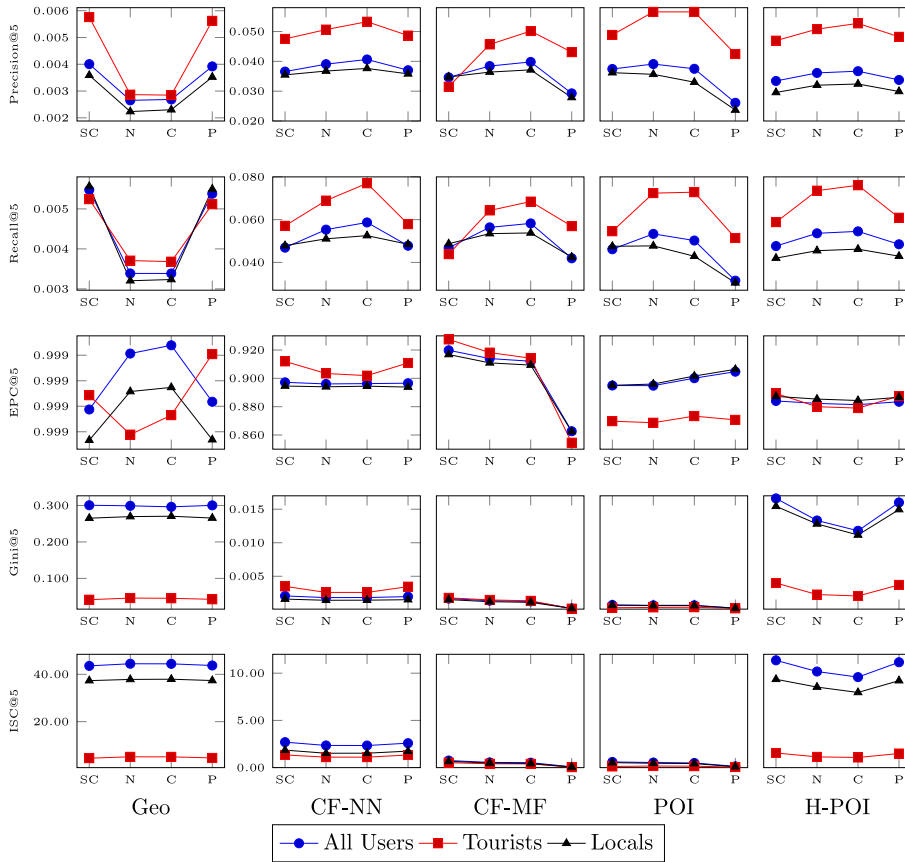


Fig. 4. Results of tourists (7.62% of the users) and local (72.43% of the users) users in Santiago; notation as in Fig. 3.

Our decision to select these three cities was because they evidence a different behavior with respect to the ratio $C_{r(C)}/C_r$ when comparing the N-MCA and C-MCA strategies: whereas TOK and SAN obtains a very similar ratio for both strategies with a percentage around 50% and 73% respectively, MEX presents substantially different values (see Table 7).

Based on these results, and considering all users, one observation that may draw our attention is that the algorithms of the Geo family have higher novelty and diversity than the other families. This is because these recommenders are based solely on recommending POIs that are close to the target user, ignoring other factors like the popularity of the POIs (something that, after observing the results obtained by the rest of the recommenders, seems to confirm the popularity bias of the data). Besides, this type of recommender is the worst in terms of accuracy, as shown by the nDCG metric in the previous figures and with Precision and Recall in the ones shown here, and it is well-known that there is typically a tradeoff between accuracy and novelty/diversity.

Regarding the other recommenders, we observe that the H-POI family tends to obtain better diversity results than the other families, although the novelty of its recommendations is usually lower. This can perhaps be explained by the fact that the best algorithm of this family is always the PGN recommender, which combines popularity and collaborative filtering with the distance between the POIs. The first two contributions reduce the novelty and diversity, but the latter, as we have seen in the Geo family, increases both dimensions so it makes sense that this approach may improve to some extent either dimensions. It is also worth considering that, in general, POI and CF-MF families achieve lower levels of novelty and diversity than the rest of the algorithms. It must be taken into account that both families use some kind of matrix factorization techniques, as the GeoBPR and IReMF. Low diversity values are indicative that very few different items are actually recommended, whereas low novelty values suggest in this case that most of the recommended POIs are those that have been visited by more users in the training set (popular items). This means that there is a significant popularity bias in the recommendations provided by these families, which is actually corroborated because their performance in terms of relevance is also high (Jannach, Lerche, Kamehkhosh, & Jugovac, 2015). On the other hand, the behavior of CF-MF in terms of EPC (novelty) might be reinforced by another aspect. In Mexico City and Santiago we observe that the novelty for the P-MCA strategy decreases steadily. We believe this might be caused because in this type of strategy, we are considerably increasing the number of items and users in the system, but at the same time we only recommend POIs from the destination city; hence, the latent factors of those POIs that are more popular are updated more frequently. If we look at the Tokyo results, the novelty of this strategy does not decrease so much because it is already very low for all strategies. This can be attributed

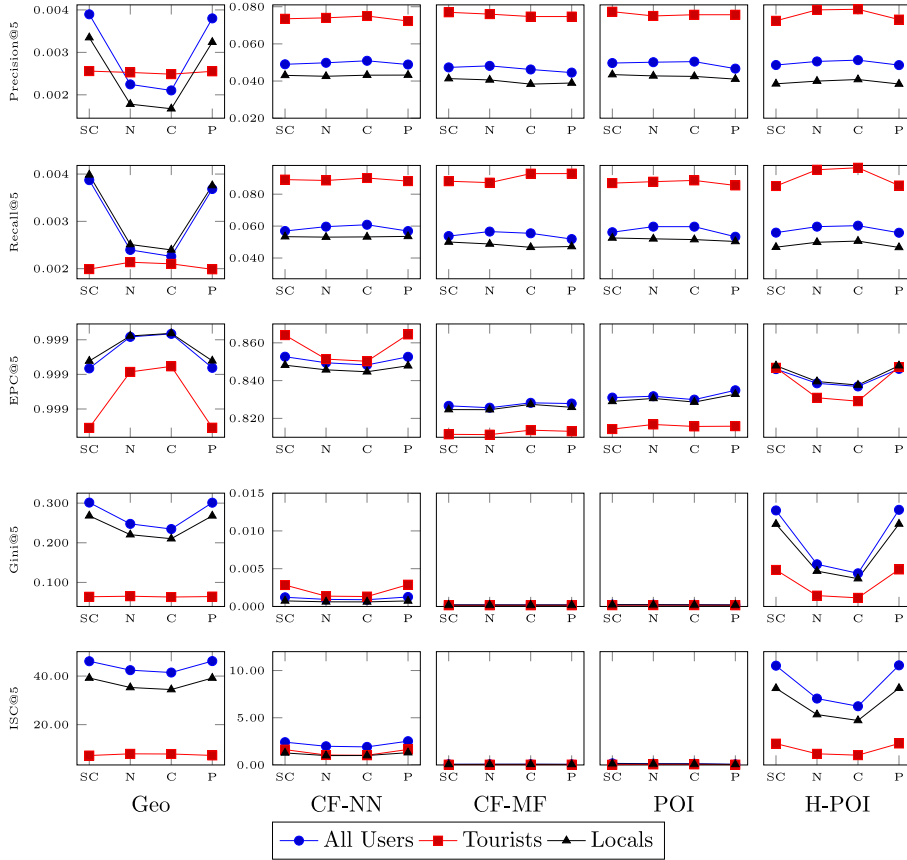


Fig. 5. Results of tourists (9.32% of the users) and local (62.44% of the users) users in Tokyo; notation as in Fig. 3.

to the fact that in Tokyo the optimal CF-MF algorithm is the BPR, which is not so sensitive to the previous behavior, while in the other two cities, is the ALS (see Table 5).

When we analyze the results by types of users (tourists or locals), it is important to consider the third type of users that is being ignored in the figure, the outlier users, together with the new users (in the test set) that do not appear in the training set and, thus, do not fit in any category. Because of this, we include in the caption of each figure the percentage of tourists and locals in the test set of the corresponding city, so the rest of the users would be labeled as outliers or no belonging to any group. This piece of information is important in this case because we show together information from all users and separated by user group. Based on these results, we corroborate that in terms of accuracy metrics (Precision and Recall) tourists achieve better performance than locals, except in the Geo family.

To better understand the rationale behind these results, we have analyzed in more detail the biases in the data and in the recommendations produced by the algorithms, and discovered that tourists are more likely to visit popular POIs (according to the training set) than locals, as evidenced by their check-ins in the test set. This is also observed in the figures, due to the higher novelty and diversity values achieved by most recommendation families for the local users (except for the CF-NN). As discussed before, this could be attributed to several reasons. Firstly, there are more local users than tourists, so it is more likely that there are more different recommended items for this type of user. Secondly, it is more likely that tourists tend to visit the most touristic venues in a city, but this should be considered in combination with the fact that locals are probably visiting a larger variety of POIs, since they spread more evenly across the city and throughout longer periods of time. This could be an explanation as to why the novelty in locals tend to be higher than in tourists.

One dimension that deserves further attention is user coverage. This measurement, as reported in Table 8, accounts for the number of users that have received at least one recommendation. In these results, the first observation we make is that the H-POI is the only recommender family with full user coverage, something that does not change when using any aggregation strategy; the reason for this is that the best recommender in this family is the PGN and one of the algorithms exploited by this hybrid recommender is based on popularity, which is a non-personalized algorithm and, hence, also has full user coverage by design. A more interesting result that emerges from this analysis is that the C-MCA strategy always improves the coverage of the recommenders, in fact, according to the column depicting the relative improvements with respect to results from the Single City, this strategy produces the largest improvements for every family in all the cities. This is a very important outcome, since together with the

Table 8

User coverage obtained by the recommenders when augmented information is used for training. The improvement in user coverage with respect to SC is represented as $\Delta(\%)$.

City	Family	SC	N	C	P	Δ N (%)	Δ C (%)	Δ P (%)
IST	Geo	81.65	83.82	87.65	81.74	2.66	7.36	0.12
	CF-NN	84.58	86.38	89.92	84.61	2.12	6.32	0.03
	CF-MF	85.10	86.93	90.25	85.19	2.14	6.05	0.10
	POI	85.10	86.93	90.25	85.19	2.14	6.05	0.10
	H-POI	100.00	100.00	100.00	100.00	0.00	0.00	0.00
JAK	Geo	86.81	91.63	92.81	87.73	5.56	6.92	1.06
	CF-NN	88.05	92.97	93.95	88.54	5.59	6.69	0.55
	CF-MF	89.51	93.49	94.51	90.43	4.45	5.59	1.03
	POI	89.51	93.49	94.51	90.43	4.45	5.59	1.03
	H-POI	100.00	100.00	100.00	100.00	0.00	0.00	0.00
KUA	Geo	80.77	88.91	91.03	81.36	10.08	12.71	0.74
	CF-NN	85.01	90.79	92.31	85.37	6.80	8.58	0.42
	CF-MF	85.30	91.06	92.56	85.95	6.76	8.51	0.77
	POI	85.30	91.06	92.56	85.95	6.76	8.51	0.77
	H-POI	100.00	100.00	100.00	100.00	0.00	0.00	0.00
MEX	Geo	88.58	90.90	93.76	88.75	2.62	5.85	0.20
	CF-NN	91.05	92.81	95.07	91.11	1.93	4.41	0.06
	CF-MF	91.27	93.04	95.34	91.46	1.94	4.46	0.21
	POI	91.27	93.04	95.34	91.46	1.94	4.46	0.21
	H-POI	100.00	100.00	100.00	100.00	0.00	0.00	0.00
MOS	Geo	85.27	85.83	89.55	85.74	0.67	5.02	0.56
	CF-NN	87.57	88.00	91.55	87.41	0.49	4.55	-0.18
	CF-MF	88.07	88.50	92.03	88.59	0.49	4.50	0.59
	POI	88.07	88.50	92.03	88.59	0.49	4.50	0.59
	H-POI	100.00	100.00	100.00	100.00	0.00	0.00	0.00
SAN	Geo	90.54	94.19	94.88	91.23	4.03	4.79	0.76
	CF-NN	92.06	95.00	95.64	92.26	3.19	3.89	0.21
	CF-MF	92.28	95.27	95.91	92.94	3.24	3.93	0.72
	POI	92.28	95.27	95.91	92.94	3.24	3.93	0.72
	H-POI	100.00	100.00	100.00	100.00	0.00	0.00	0.00
SAO	Geo	82.21	85.81	90.44	82.62	4.37	10.01	0.50
	CF-NN	83.33	87.84	92.02	83.46	5.41	10.43	0.15
	CF-MF	85.04	88.51	92.69	85.37	4.08	8.99	0.39
	POI	85.04	88.51	92.69	85.37	4.08	8.99	0.39
	H-POI	100.00	100.00	100.00	100.00	0.00	0.00	0.00
TOK	Geo	87.75	92.68	93.64	88.01	5.61	6.71	0.30
	CF-NN	90.32	94.15	94.78	90.32	4.24	4.94	0.00
	CF-MF	90.52	94.29	94.96	90.78	4.17	4.91	0.29
	POI	90.52	94.29	94.96	90.78	4.17	4.91	0.29
	H-POI	100.00	100.00	100.00	100.00	0.00	0.00	0.00

results shown in Figs. 3, 4 and 5, where this strategy obtains better or equal accuracy results than N-MCA for most of the families in every city, it means that it is able to improve the results for more users in the system, simply by integrating carefully selected additional information. This process, in any case, would be achieved at a lower cost than the P-MCA strategy, hence, allowing more efficient computations. In particular, this allows us to create a single training set containing the check-ins of the cities we need and make recommendations from this training, instead of making an independent set by each city and hence allowing more efficient computations. Nevertheless, we can observe an interesting result in the city of Moscow. If we analyze the user coverage of this city more in detail, we can see how the user coverage decreases in the P-MCA strategy for CF-NN recommenders. This behavior, although it may seem counter intuitive, is due to the fact that those algorithms that work with similarities between users or items (users, in this case) might obtain a large number of neighbors with the same degree of similarity, some of them coming from the aggregated cities and with potentially no check-ins in the target city. In this sense, and according to these results, by using the P-MCA strategy it is more likely that we may find new neighbors with near-zero overlap with items in the target city that are, hence, not able to recommend any POI there, thus, reducing the coverage of such algorithms.

Finally, we can provide an answer for RQ3 (*How do MCA strategies affect other evaluation dimensions? What is the impact of these dimensions on different groups of users?*). In general, we have observed that the N-MCA strategy is the safest one both in terms of accuracy and beyond-accuracy metrics, although C-MCA obtains very similar results while improving the user coverage, hence, impacting positively to more users. These strategies also show good results for tourists, although all users get some kind of improvement with these approaches. Regarding the effect in the groups of users, tourists are positively affected in terms of accuracy but negatively for other dimensions such as novelty and diversity.

Table 9
Performance in terms of nDCG@5 of the Popularity recommender in all cities in both Tourists and Locals.

City	All Users	Tourists	Locals	Δ Tourists (%)	Δ Locals (%)
Istanbul	0.054	0.064	0.048	19.04	-9.77
Jakarta	0.066	0.091	0.053	38.33	-19.92
Kuala Lumpur	0.066	0.077	0.060	17.34	-8.46
Mexico City	0.041	0.059	0.034	45.69	-15.70
Moscow	0.027	0.037	0.026	34.02	-4.48
Santiago	0.051	0.067	0.044	30.47	-13.21
São Paulo	0.053	0.061	0.031	14.85	-40.33
Tokyo	0.069	0.106	0.056	53.48	-18.73

5.4. Discussion

According to the presented results, applying Multi-City Aggregation strategies to augment the data available in venue recommendation can improve the results obtained in some situations, although their effect is not as great as one might expect (mainly due to the temporal split we used and the dataset being too sparse). Nevertheless, since we explored some basic recommendation techniques across a wide range of algorithmic families, these results are promising and may open the door to debate about the importance of the geographical distribution of the check-ins in evaluation. First of all, because we have seen that some algorithms are able to make better recommendations (in some cases, up to a 20% improvement), and in some situations – mostly under the C-MCA strategy – the user coverage is enhanced; however, further analysis should be done to properly understand the impact of such improvements in other evaluation dimensions, such as novelty or diversity, and how it generalizes to different cities. Similarly, the N-MCA strategy has generally returned cities belonging to the same country with respect to the target city (except in the case of Santiago, where one of its 7 closest cities was Córdoba, a city of Argentina). This opens up the possibility that cities different from those used in this paper may have nearby cities in other countries. Therefore, it would also be interesting to further analyze the differences between C-MCA and N-MCA strategies in other cities.

Secondly, due to the well-known popularity bias (Jannach et al., 2015), such a simple technique could outperform other methods like IB, KDE, or AvgDis (see Table 5), and it has resulted in a very positive component to be integrated in a hybrid algorithm (i.e., PGN), even though this type of baseline is usually ignored in POI recommendation literature. Thirdly, we have observed that it is not the same to train the recommenders with interactions of a certain city as training them with the check-ins of a whole country. Hence, POI recommendation proposals that are trained using information of specific regions may not be comparable to others trained with data from around the world. We want to emphasize that in our work we have only used cross-domain techniques oriented at exploiting user information (i.e., by maximizing user overlap). As indicated in Section 5.1, we leave as future work other cross-domain techniques that could be more appropriate for item similarity models.

At the same time, we have observed that if we distinguish between two types of users (tourists and locals), almost every recommendation algorithm produces very different results to each user group. To further understand this, we now analyze in more detail the effect of popularity bias in our experimental settings. First, in Table 9 we show the results of the Popularity recommender in the Single City (SC) configuration. In that table we show the results obtained taking into account all users in the test set as well as the results for the tourists and locals. The last two columns represent the change in performance (as a percentage, negative or positive depending on whether the performance improved or decreased) obtained by each group of users with respect to the value obtained by considering that all users belong to the same group. As we can observe, tourists tend to obtain a result between a 14% and a 50% higher than the rest of the users. This seems to confirm that tourist users tend to visit more popular POIs than locals. To further view this effect in detail, in Fig. 6 we show the top 1% of the most popular POIs of every city that appear in the test set with the percentage of users (of each group) that checked-in in that POI. In this figure we can see how the most popular POIs in general receive more visits from tourist users (relatively) than from local users. This makes sense since, as we have indicated before, for a user to be considered a tourist she has to perform check-ins in the city for at most 21 days, so it is more likely that many of those tourists did not have enough time to visit the most popular POIs in the training set and hence they visit them in the test set.

Nevertheless, it is well-known that, besides popularity bias, there might be other biases in the data (Boratto, Fenu, & Marras, 2019). In particular, in LBSN it is common to associate each POI with one or more categories (restaurant, museum, park, bar, etc.). While we can establish different category hierarchies in Foursquare, we will focus on level 1 categories, since the number of categories (9) is more manageable than other, more specific levels. Using these categories, in Fig. 7 we show the percentage of check-ins of every user group in each category of the items. From this figure, even though it is noticeable a slightly different behavior of tourists with respect to the rest of the users, we do not observe major differences, except perhaps in the cities of Santiago, São Paulo, and Tokyo. This indicates that, despite an overall category bias exists (since POIs related to food and transportation receive more check-ins than the rest of POIs), there is no difference in behavior between the different groups of users in terms of the categories they visit. Hence, together with our previous observation, this shows that tourist and local users tend to go to the same type of POIs, but tourist users tend to choose to visit more popular venues (of the same categories) than local users.

These results evidence a general trend or systematic bias, since it is much easier to make relevant recommendations to tourists due to the type of POIs they usually visit (those already popular in a city). In particular, this observation – which is, to the best of our knowledge, novel in the area – would open up several possibilities in terms of deciding how many resources should be devoted to each user group. Additionally, a negative result we observed is that when the distance between the venues is considered in the

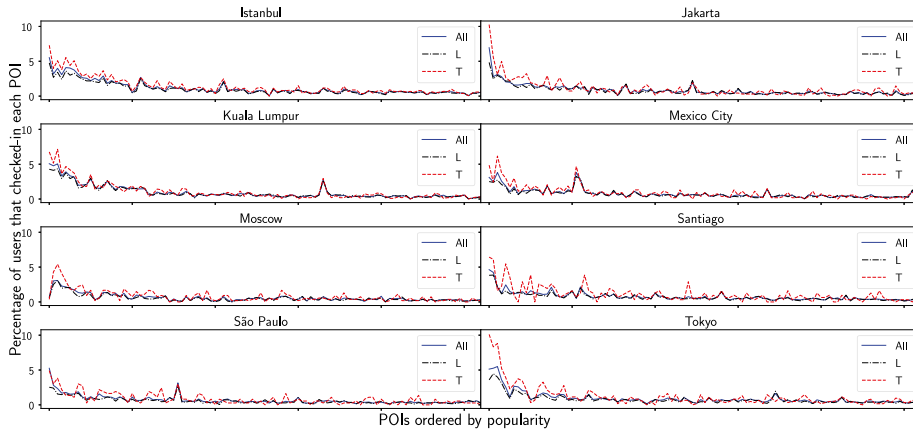


Fig. 6. Popularity bias in the eight selected cities. Each plot shows the percentage of users belonging to each group who have a check-in in the test set for each corresponding item. Items are sorted according to their popularity in the training set.

recommendation algorithm, augmenting the available information through MCA strategies can be counterproductive in terms of accuracy, although other dimensions might benefit from such augmentation. This effect is not conclusive for the two types of users considered, although we have observed a negative trend for tourists, whose diversity and novelty values tend to be much lower than those for local users, in particular when MCA strategies are exploited.

It should be noted that our results are consistent with those discussed in [Sahebi and Brusilovsky \(2015\)](#). The authors found that there are specific experiments where cross-domain recommendation works worse than classic recommendation, even though in general it behaves better or as good as strategies where cross-domain is not used (single-domain). However, only comparisons between single- and cross-domain approaches on three different algorithms and without considering any temporal split were presented in that study; hence, our work helps on generalizing the conclusions obtained in such paper.

We want to emphasize that considering information from different cities (understood as different domains), despite being computationally more expensive, has a clear advantage: such system would only need to train once whenever recommendations are required for any of the cities included in the MCA strategy, whereas considering each city as an isolated training domain (when no MCA strategy is used) only allows to generate recommendations for a single city; hence, the recommendation model built in such a way can be re-used more often in the former case, at the expense of being more expensive (although this would depend on the actual strategy considered) in terms of memory and time consumption. This conclusion may help other researchers in the area in order to apply a more favorable data preprocessing for the POI recommendation models that they are developing. Nonetheless, as we have shown here, if the MCA strategy is generated based on the *right* cities, significant performance improvements can be achieved, not always by selecting the cities with more information but those that are closer and more likely to have overlap in their users, probably because they are culturally related and share similar mobility patterns ([Yang et al., 2016](#)).

6. Related work

6.1. Data augmentation and cross-domain recommendation

As already discussed before, extending the available information with additional data can help recommendation systems in a number of situations. Specifically, travel recommendation was addressed as a potential target of these techniques ([Zheng, 2015](#)). There have been some papers in which researchers explored different ways to combine sources of information to be applied in tourism. For example, in [Sabou, Onder, Brasoveanu, and Scharl \(2016\)](#) the authors describe TourMIS, a dataset of European statistics of tourism data, where they show the usefulness of combining different data sources (economy, tourism, and sustainability) to make relations between them. Although they indicate that sometimes it is difficult to integrate sources from very different domains, most of those problems can be solved using Linked Data approaches. Nevertheless, the dataset described was not used to produce venue recommendations, only for statistical analysis.

A particular instance of those techniques that augment the available information are the so-called cross-domain recommendation algorithms ([Cantador et al., 2015](#)): the basic idea behind this type of recommendation is that, to improve recommendations over a target domain, some kind of knowledge from a source domain needs to be exploited. However, we have not found many examples of cross-domain experiments combining more than two or three domains — usually, movies, music, and books, except for the works presented in [Rafailidis and Crestani \(2017\)](#) and [Sahebi and Brusilovsky \(2015\)](#). In the first one, the authors exploited the information of 10 domains (categories of different products from Epinions, hence, not related with tourism) in order to analyze the performance of the recommendations using ranking metrics (nDCG and Recall), however, these datasets are much smaller in terms of ratings than the ones we use in this paper (the largest one contains around 200K ratings) and they are all more dense. Similarly, in [Yu et al. \(2018\)](#), the authors propose a cross-domain algorithm that combines two source domains and test its performance for only 586

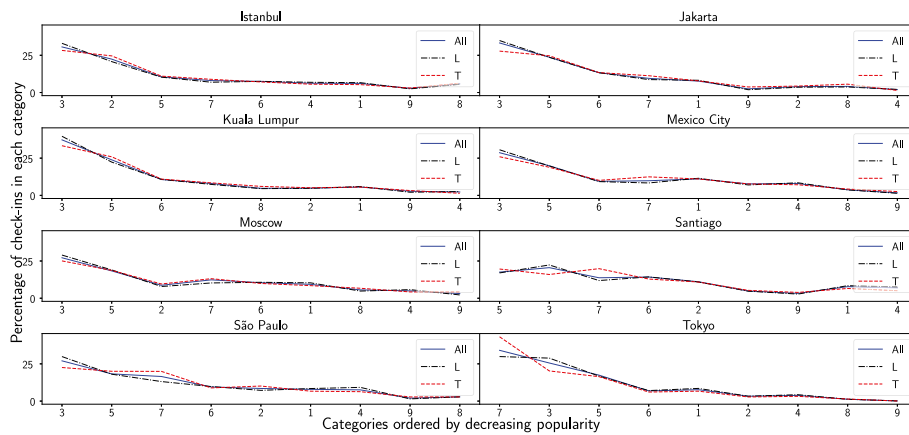


Fig. 7. Popularity bias at category-level in the eight selected cities. Each plot shows the percentage of interactions a user group performed in test to items belonging to each category. Categories are ordered by decreasing popularity in each city. The numbers correspond to their Foursquare ids: (1) Arts & Entertainment, (2) Outdoors & Recreation, (3) Food, (4) Nightlife Spots, (5) Shops & Services, (6) Professional & Other Places, (7) Travel & Transport, (8) Colleges & Universities, and (9) Residences.

users out of 1.5M users from the original Amazon dataset. Finally, in [Sahebi and Brusilovsky \(2015\)](#), the authors used 21 domains (defined as the categories from different Yelp businesses) and measured the effects of cross-domain recommendation in terms of RMSE by comparing the performance on several domain pairs; however, the interest on error-based metrics such as RMSE from the RS community has decreased since they do not correlate with user satisfaction at the same level as ranking-based metrics ([Cremonesi, Garzotto, Negro, Papadopoulos, & Turrin, 2011](#); [McNee, Riedl, & Konstan, 2006](#)). Therefore, our work is – to best of our knowledge – the first study where up to 8 domains are combined and used as the source domain (since each city can be interpreted as a separate domain), and in the context of a tourism dataset with a realistic, temporal evaluation focused on ranking-based metrics.

To complement these works, one approach of cross-domain in tourism can be seen in [Fernández-Tobías, Cantador, Kaminskas, and Ricci \(2011\)](#), where the authors did not recommend POIs but music artists depending on the monument the user is visiting; they do this by building a semantic network of venues and artists and the relations between them. Moreover, in [Zheng, Zheng, Xie, and Yang \(2010\)](#), the authors exploited information from different sources like location–activity, location–feature, and activity–activity correlations to enhance the performance, showing a 20% improvement over a basic algorithm that does not use any additional data. Additionally, in [Manotumruksa, Rafailidis, Macdonald, and Ounis \(2019\)](#) the authors analyze the impact of CrossFire, a cross-domain framework in venue recommendation, against CF-MF models and a deep recurrent collaborative filtering model. According to their results, the cross-domain framework does not clearly contribute to an improvement in these recommendations algorithms.

Nonetheless, the most similar approach to this paper that we found is [Zhang and Wang \(2016\)](#). In that article, the authors perform a so-called cross-region recommendation, and considered each region as a different domain. One important difference with our work is that, whereas a division by cities is natural, that work uses regions computed by performing clustering on the venues. Additionally, the datasets used (Foursquare and Yelp) are smaller than the one reported here and no temporal evaluation was performed, only a standard cross validation methodology. Hence, our paper offers a complementary view on a related problem, from a more realistic perspective (since we explicitly address a time-aware evaluation) with a larger dataset and taking into account the check-ins in different cities to perform the recommendations.

6.2. Biases and types of users

At the same time, thanks to the heterogeneous and ever-growing data available in the Recommender Systems, it is possible – as we have already discussed – to establish different user groups depending on available user features (such as gender, nationality, age, etc.), the distribution of ratings or other data ([Deldjoo, Anelli, Zamani, Bellogín, & Noia, 2021](#); [Sánchez & Bellogín, 2019](#)). The analysis of these groups is useful in order to detect possible biases that the recommenders may inherently have, reproduce, or reinforce because of the data. For example, in [Ekstrand, Tian, Kazi, Mehrpouyan, and Kluver \(2018\)](#) the authors analyze the recommendations produced in the book domain (in the Amazon and BookCrossing datasets) by collaborative filtering algorithms according to the gender of the authors, concluding that matrix factorization algorithms tend to have biased recommendations towards male authors, which corresponds to the most frequent group in their data, hence neglecting the less representative part of the population when learning preferences from the global population. On the other hand, in [Abdollahpouri, Mansoury, Burke, and Mobasher \(2019\)](#), the authors analyze the effect of the well-known popularity bias in the MovieLens dataset by characterizing the users in three different groups according to the percentage of popular items they have rated. According to their results, all algorithms recommended many more popular items than the ones the users have rated, even for the group of users in which half of the items they consumed belonged to the long-tail.

In the area of tourism and POI recommendation, the analysis of biases and different types of users is even more critical due to the sparsity of the data and the nature of the items considered. Moreover, this information could be very useful to understand the

actual behavior of users, either as a whole or as members of smaller groups. In this regard, in [Dietz et al. \(2019\)](#) the authors apply a clustering algorithm to obtain 4 major user groups (vacationers, explorers, voyagers, and globetrotters) based on the types of trips they make in a Foursquare dataset. In [Bao et al. \(2012\)](#) a POI recommendation approach is proposed that takes into consideration the local experts in the cities in order to improve the recommendations, but these were focused to only foreign travelers. Additionally, in [Neidhardt, Seyfang, Schuster, and Werthner \(2015\)](#) the authors define a Seven Factor Model of touristic patterns in which they capture the user preferences by letting them choose the most interesting photos for them from a collection of images. Later on, both users and items are transformed into a 7-dimensional space (one for each factor) so that the recommendation for the user is computed applying the Euclidean distance between the user and the item transformations. In this way, hence, an alternative method for user profiling in the tourism domain, in this case considering the emotional information encoded in the pictures, is proposed.

While the aforementioned works have explored different techniques to group or classify users in LBSNs, we have not found many works where the effect of the recommendation algorithms has been explicitly analyzed towards each user group. A related, but different problem, nonetheless, is addressed in [Wattanacharoensil and La-or-nual \(2019\)](#), where the authors show that cognitive biases such as confirmation, positive (or negative), stereotype, temporal perspective, and so on affect the decisions of the tourists. Another approach can be found in [Weydemann, Sacharidis, and Werthner \(2019\)](#), where the authors analyze different fairness criteria in location recommendations in the city of Vienna considering groups of users based on their nationalities. We, thus, believe this work would allow to better understand how to design these types of systems and to consider such biases when modeling the users.

7. Conclusions

We have explored the venue recommendation problem and compared the performance of several algorithms under a realistic scenario using a temporal evaluation methodology. Our main contribution consists in different ways of augmenting the information available to train recommendation algorithms. More specifically, we have shown an empirical evaluation comparing the performance of state-of-the-art recommenders under different data augmentation settings (two aggregation strategies based on the closest cities, with the special case of the cities belonging to the same country, and the most popular ones). Even though the behavior varies depending on the city, the data augmentation strategy based on the closest cities (and, in particular, the one based on the cities of the same country) tends to produce better results and, what is more important, in more than one evaluation dimension; in the future, we would like to exploit this effect to create a generic recommender for venue suggestion that takes this information into account.

Encouraged by these positive results, we believe there is still room for improvement. First of all, this study could be extended by considering more complex recommendation algorithms such as LTR ([Gao et al., 2013](#)) or others based on neural networks ([Manotumruksa, Macdonald, & Ounis, 2017](#); [Tang & Wang, 2018](#)); although we should consider that POI recommendation models that exploit geographic information may be negatively affected by the MCA strategies as user movement patterns may be modified if we use information from other cities' check-ins. Furthermore, the temporal evaluation methodology should also be analyzed more carefully, especially regarding the effect of seasonal trends and how it may affect the data augmentation and knowledge transfer techniques (since not enough interactions or users might be available or active at the same time). Nevertheless, we still argue that this type of offline evaluation is more realistic as it resembles a real production system where only past interactions are available to predict future behavior, unlike random splits where past and future interactions are mixed ([Zhang & Chow, 2015a](#)).

Another contribution that we have analyzed in this work consists on identifying the systematic biases that groups of users from LBSNs experience when receiving recommendations. We have focused on two clearly different types of users: tourists and locals. Our results indicate that state-of-the-art venue recommenders consistently find better suggestions for tourists, which seems to demonstrate that these users are easier to satisfy. This behavior is also observed for the data augmentation techniques proposed, although it is strongly connected to measuring performance based on accuracy, since other dimensions such as diversity or novelty show an inverse relation with accuracy and, hence, better results for local users.

These results open up interesting research lines for the future. On the one hand, it suggests that the types of users analyzed in this work (tourists and locals) could be optimized according to different criteria or even using different pools of algorithms. By doing this, more resources could be devoted to those users more difficult to optimize (in this case, local users), an idea that was explored in the area and proved to decrease the total error in the system ([Saïd & Bellogín, 2018](#)). Another possibility would be to create hybrid recommendation algorithms that, depending on the user type, would exploit different information to provide suggestions. We believe these possibilities would allow to decrease the biases observed herein, since they would allow to provide better recommendations for the population whose preferences are being misrepresented.

Furthermore, we aim to further extend how the cities are selected by computing a similarity between them using content and cultural information ([Yang et al., 2016](#)), and in particular, item-based similarities that could reduce the sparsity when aggregating the cities by merging POIs that belong to the same food or clothing chain, as done in [Karamshuk, Noulas, Scellato, Nicosia, and Mascolo \(2013\)](#). We would also like to exploit the venue categories, social connections, or other content data, and see how that information is affected when using aggregation strategies, especially in the context of smart city applications where multiple data sources are available [Lau et al. \(2019\)](#). Besides, due to the inherent high sparsity levels of this recommendation task, we believe that a study to analyze the cold-start problem is needed in order to assess if data augmentation strategies perform well under these circumstances. In fact, since in this paper we have found that the proposed strategies are useful to improve the user coverage even under very sparse constraints, we expect they could obtain positive results also for such a problem, in particular because our experimental setting is still realistic enough under several situations (we applied a 2-core pre-processing step, which is less strict

than other settings used in the area); nonetheless, a more thorough analysis such as the one reported in [Kluver and Konstan \(2014\)](#) should be performed to understand the sensitivity of this problem to the different algorithms and aggregation strategies.

Finally, we plan to expand the analysis on biases to the types of users proposed in [Dietz et al. \(2019\)](#), which considered different types of travelers (and, hence, they would all classify as tourists under our classification). Moreover, we believe there is room for improvement in terms of grouping the users according to different dimensions, and we would like to explore other classifications, for instance, based on the number of interactions or on the number of popular POIs they visit, as done in recent works ([Abdollahpouri et al., 2019](#); [Sánchez & Bellogín, 2019](#)).

CRedit authorship contribution statement

Pablo Sánchez: Conceptualization, Methodology, Software, Validation, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. **Alejandro Bellogín:** Conceptualization, Methodology, Formal analysis, Supervision, Writing - original draft, Writing - review & editing.

Acknowledgments

This work has been co-funded by the European Social Fund (ESF) within the 2017 call for predoctoral contracts, the Ministerio de Ciencia e Innovación (ref: PID2019-108965GB-I00), and by the project “AmerMad-CM: America in Madrid. Interconnected heritage and impact on tourism in the Community of Madrid” (ref: H2019/HUM-5694), co-financed by the Community of Madrid and the European Social Fund (ESF).

Appendix. Information about N-MCA and C-MCA cities

Here we include the 7 closest cities (N-MCA strategy) with respect to each target city:

- Istanbul: Kutahya, Bursa, Eskisehir, Tekirdag, Kocaeli, Balikesir, Sakarya.
- Jakarta: Palembang, Tanjungkarang-Telukbetung, Pontianak, Bandung, Surabaya, Semarang, Yogyakarta.
- Kuala Lumpur: Ipoh, Seremban, Pinang, Kuantan New Port, Shah Alam, Kuala Terengganu, Melaka.
- Mexico City: Queretaro, Jalapa, Morelia, Puebla, Pachuca, Toluca, Cuernavaca.
- Moscow: Tver, Yaroslavl, Ivanovo, Gor'kiy, Voronezh, Ceboksary, Smolensk.
- Santiago: Valparaíso, Coquimbo, Talca, Córdoba, Concepción, Temuco, La Serena.
- São Paulo: Florianópolis, Curitiba, Santos, Vitoria, Belo Horizonte, Niteroi, Rio de Janeiro.
- Tokyo: Sendai, Kawasaki, Osaka, Gifu, Yokohama, Kyoto, Nagoya.

To properly compare the results from N-MCA and C-MCA, we also state all cities belonging to the same country (C-MCA) with respect to each target city:

- Turkey: Istanbul, Sakarya, Balikesir, Canakkale, Zonguldak, Kutahya, Kayseri, Ordu, Trabzon, Mersin, Isparta, Mugla, Denizli, Sanhurfa, Aydin, Ankara, Eskisehir, Malatya, Kocaeli, Seyhan, Tekirdag, Afyon, Samsun, Rize, Izmir, Bursa, Antalya, Giresun, Antioch, Manisa, Kahramanmaras, Bolu, Edirne, Konya, Aintab.
- Indonesia: Jakarta, Palembang, Tanjungkarang-Telukbetung, Semarang, Samarinda, Balikpapan, Surabaya, Bandung, Denpasar, Bandjermasin, Mataram, Yogyakarta, Padang, Pontianak, Medan, Manado, BandaAceh, Pekanbaru.
- Myanmar: Kuala Lumpur Ipoh, Alor Setar, Melaka, Kangar, Kuantan NewPort, Pinang, Kota Baharu, Kuala Terengganu, Kota Kinabalu, Kuching, Johor Baharu, Seremban, ShahAlam.
- Mexico: Mexico City, Villahermosa, Queretaro, Tampico, Jalapa, Morelia, Puebla, Pachuca, Toluca, Cuernavaca, Guadalajara, Aguascalientes, La Paz, Oaxaca, Tuxtla Gutiérrez, San Luis Potosí, Campeche, Colima, Veracruz, Mérida, Monterrey, Hermosillo.
- Russia: Moscow, Irkutsk, Kuybyshev, Kaliningrad, Ivanovo, Voronezh, Vladivostok, Gor'kiy, Chelyabinsk, Rostov-on-Don, Omsk, Krasnodar, Perm, Novosibirsk, Vyatka, Saint Petersburg, Tver, Ufa, Tomsk, Smolensk, Sverdlovsk, Krasnoyarsk, Volgograd, Kazan, Izevsk, Ceboksary, Ulyanovsk, Yakutsk, Khabarovsk, Yaroslavl, Saratov.
- Chile: Santiago, Puerto Montt, Valparaíso, Coquimbo, Talca, Antofagasta, Concepción, Temuco, La Serena, Iquique.
- Brazil: São Paulo, Joao Pessoa, Porto Velho, Natal, Palmas, Belem, Manaus, Maceio, Aracaju, Boa Vista, Vitoria, Niteroi, Brasilia, Belo Horizonte, Cuiaba, Sao Luis, Macapa, Curitiba, Rio de Janeiro, Rio Branco, Goiania, Florianópolis, Teresina, Fortaleza, Santos, Campo Grande, Recife, Porto Alegre, Santarem, Salvador.
- Japan: Tokyo, Hiroshima, Naha, Fukuoka, Kobe, Kawasaki, Sendai, Kawasaki, Osaka, Gifu, Yokohama, Kyoto, Nagoya, Shimonoseki, Sapporo.

References

- Abdollahpouri, H., Mansoury, M., Burke, R., & Mobasher, B. (2019). The unfairness of popularity bias in recommendation. In R. Burke, H. Abdollahpouri, E. C. Malthouse, K. P. Thai, & Y. Zhang (Eds.), *CEUR workshop proceedings: Vol. 2440, Proceedings of the workshop on recommendation in multi-stakeholder environments co-located with the 13th ACM conference on recommender systems*. CEUR-WS.org.
- Aioli, F. (2013). Efficient top-n recommendation for very large scale binary rated datasets. In Q. Yang, I. King, Q. Li, P. Pu, & G. Karypis (Eds.), *Seventh ACM conference on recommender systems* (pp. 273–280). ACM.
- Bao, J., Zheng, Y., & Mokbel, M. F. (2012). Location-based and preference-aware recommendation using sparse geo-social networking data. In I. F. Cruz, C. A. Knoblock, P. Kröger, E. Tanin, & P. Widmayer (Eds.), *SIGSPATIAL 2012 international conference on advances in geographic information systems (Formerly Known As GIS)* (pp. 199–208). ACM.
- Bao, J., Zheng, Y., Wilkie, D., & Mokbel, M. F. (2015). Recommendations in location-based social networks: a survey. *Geoinformatica*, 19(3), 525–565.
- Boratto, L., Fenu, G., & Marras, M. (2019). The effect of algorithmic bias on recommender systems for massive open online courses. In L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff, & D. Hiemstra (Eds.), *Lecture notes in computer science: Vol. 11437, Advances in information retrieval - 41st European conference on IR research, ECIR 2019, Cologne, Germany, April 14–18, 2019, proceedings, Part I* (pp. 457–472). Springer.
- Cantador, I., Fernández-Tobías, I., Berkovsky, S., & Cremonesi, P. (2015). Cross-domain recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 919–959). Springer.
- Castells, P., Hurley, N. J., & Vargas, S. (2015). Novelty and diversity in recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 881–918). Springer.
- Cheng, C., Yang, H., King, I., & Lyu, M. R. (2012). Fused matrix factorization with geographical and social influence in location-based social networks. In J. Hoffmann, & B. Selman (Eds.), *Proceedings of the twenty-sixth AAAI conference on artificial intelligence*. AAAI Press.
- Chiang, H., & Huang, T. (2015). User-adapted travel planning system for personalized schedule recommendation. *Information Fusion*, 21, 3–17.
- Choudhury, M. D., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., & Yu, C. (2010). Automatic construction of travel itineraries using social breadcrumbs. In M. H. Chignell, & E. G. Toms (Eds.), *HT'10, proceedings of the 21st ACM conference on hypertext and hypermedia* (pp. 35–44). ACM.
- Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A. V., & Turrin, R. (2011). Comparative evaluation of recommender system quality. In D. S. Tan, S. Amershi, B. Begole, W. A. Kellogg, & M. Tungare (Eds.), *Proceedings of the international conference on human factors in computing systems* (pp. 1927–1932). ACM.
- de Gemmis, M., Lops, P., Musto, C., Narducci, F., & Semeraro, G. (2015). Semantics-aware content-based recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 119–159). Springer.
- Deldjoo, Y., Anelli, V. W., Zamani, H., Bellogín, A., & Noia, T. D. (2021). A flexible framework for evaluating user and item fairness in recommender systems. *User Model. User Adapt. Interact.*, to appear.
- Dietz, L. W., Roy, R., & Wörndl, W. (2019). Characterisation of traveller types using check-in data from location-based social networks. In J. Pesonen, & J. Neidhardt (Eds.), *Information and communication technologies in tourism 2019, ENTER 2019, proceedings of the international conference in Nicosia, Cyprus* (pp. 15–26). Springer.
- Edizel, B., Bonchi, F., Hajian, S., Panisson, A., & Tassa, T. (2020). Fairecsys: mitigating algorithmic bias in recommender systems. *International Journal of Data Science Analysis*, 9(2), 197–213.
- Ekstrand, M. D., Tian, M., Azpiazu, I. M., Ekstrand, J. D., Anuyah, O., McNeill, D., et al. (2018). All the cool kids, how do they fit in? Popularity and demographic biases in recommender evaluation and effectiveness. In S. A. Friedler, C. Wilson (Eds.), *Proceedings of machine learning research: Vol. 81, Conference on fairness, accountability and transparency* (pp. 172–186). PMLR.
- Ekstrand, M. D., Tian, M., Kazi, M. R. I., Mehropouyan, H., & Kluver, D. (2018). Exploring author gender in book rating and recommendation. In S. Pera, M. D. Ekstrand, X. Amatriain, & J. O'Donovan (Eds.), *Proceedings of the 12th ACM conference on recommender systems* (pp. 242–250). ACM.
- Fernández-Tobías, I., Cantador, I., Kaminskas, M., & Ricci, F. (2011). A generic semantic-based framework for cross-domain recommendation. In *Proceedings of the 2nd international workshop on information heterogeneity and fusion in recommender systems* (pp. 25–32). New York, NY, USA: ACM.
- Gao, H., Tang, J., Hu, X., & Liu, H. (2013). Exploring temporal effects for location recommendation on location-based social networks. In Q. Yang, I. King, Q. Li, P. Pu, & G. Karypis (Eds.), *Seventh ACM conference on recommender systems* (pp. 93–100). ACM.
- Gao, H., Tang, J., & Liu, H. (2015). Addressing the cold-start problem in location recommendation using geo-social correlations. *Data Mining and Knowledge Discovery*, 29(2), 299–323.
- Gunawardana, A., & Shani, G. (2015). Evaluating recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 265–308). Springer.
- Han, J., & Yamana, H. (2020). Geographic diversification of recommended POIs in frequently visited areas. *ACM Transactions on Information Systems*, 38(1), 1:1–1:39.
- Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In *Proceedings of the 8th IEEE international conference on data mining* (pp. 263–272). IEEE Computer Society.
- Jannach, D., Lерche, L., Kamekhosh, I., & Jugovac, M. (2015). What recommenders recommend: an analysis of recommendation biases and possible countermeasures. *User Modeling and User-Adapted Interaction*, 25(5), 427–491.
- Kabbur, S., Ning, X., & Karypis, G. (2013). FISM: factored item similarity models for top-n recommender systems. In I. S. Dhillon, Y. Koren, R. Ghani, T. E. Senator, P. Bradley, R. Parekh, J. He, R. L. Grossman, & R. Uthrusamy (Eds.), *The 19th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 659–667). ACM.
- Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., & Mascolo, C. (2013). Geo-spotting: mining online location-based services for optimal retail store placement. In I. S. Dhillon, Y. Koren, R. Ghani, T. E. Senator, P. Bradley, R. Parekh, J. He, R. L. Grossman, & R. Uthrusamy (Eds.), *The 19th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 793–801). ACM.
- Kluver, D., & Konstan, J. A. (2014). Evaluating recommender behavior for new users. In A. Kobsa, M. X. Zhou, M. Ester, & Y. Koren (Eds.), *Eighth ACM conference on recommender systems* (pp. 121–128). ACM.
- Koren, Y., & Bell, R. M. (2015). Advances in collaborative filtering. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 77–118). Springer.
- Koren, Y., Bell, R. M., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8), 30–37.
- Lau, B. P. L., Hasala, M. S., Zhou, Y., Hassan, N. U., Yuen, C., Zhang, M., et al. (2019). A survey of data fusion in smart city applications. *Information Fusion*, 52, 357–374.
- Levandovski, J. J., Sarwat, M., Eldawy, A., & Mokbel, M. F. (2012). LARS: a location-aware recommender system. In A. Kementsietsidis, & M. A. V. Salles (Eds.), *IEEE 28th international conference on data engineering* (pp. 450–461). IEEE Computer Society.
- Li, X., Cong, G., Li, X., Pham, T. N., & Krishnaswamy, S. (2015). Rank-geofm: A ranking based geographical factorization method for point of interest recommendation. In R. A. Baeza-Yates, M. Lalmas, A. Moffat, & B. A. Ribeiro-Neto (Eds.), *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (pp. 433–442). ACM.
- Lian, D., Ge, Y., Zhang, F., Yuan, N. J., Xie, X., Zhou, T., et al. (2018). Scalable content-aware collaborative filtering for location recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 30(6), 1122–1135.

- Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., & Rui, Y. (2014). Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In S. A. Macskassy, C. Perlich, J. Leskovec, W. Wang, & R. Ghani (Eds.), *The 20th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 831–840). ACM.
- Liu, Y., Pham, T., Cong, G., & Yuan, Q. (2017). An experimental evaluation of point-of-interest recommendation in location-based social networks. *PVLDB*, 10(10), 1010–1021.
- Liu, Y., Wei, W., Sun, A., & Miao, C. (2014). Exploiting geographical neighborhood characteristics for location recommendation. In J. Li, X. S. Wang, M. N. Garofalakis, I. Soboroff, T. Suel, & M. Wang (Eds.), *Proceedings of the 23rd ACM international conference on conference on information and knowledge management* (pp. 739–748). ACM.
- Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., & Zhang, G. (2015). Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems*, 80, 14–23.
- Manotumruksa, J., Macdonald, C., & Ounis, I. (2017). A deep recurrent collaborative filtering framework for venue recommendation. In E. Lim, M. Winslett, M. Sanderson, A. W. Fu, J. Sun, J. S. Culpepper, E. Lo, J. C. Ho, D. Donato, R. Agrawal, Y. Zheng, C. Castillo, A. Sun, V. S. Tseng, & C. Li (Eds.), *Proceedings of the 2017 ACM on conference on information and knowledge management* (pp. 1429–1438). ACM.
- Manotumruksa, J., Macdonald, C., & Ounis, I. (2019). A contextual recurrent collaborative filtering framework for modelling sequences of venue checkins. *Information Processing and Management*, Article 102092.
- Manotumruksa, J., Rafailidis, D., Macdonald, C., & Ounis, I. (2019). On cross-domain transfer in venue recommendation. In L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff, & D. Hiemstra (Eds.), *Lecture notes in computer science: Vol. 11437, Advances in information retrieval - 41st European conference on IR research* (pp. 443–456). Springer.
- McNee, S. M., Riedl, J., & Konstan, J. A. (2006). Being accurate is not enough: how accuracy metrics have hurt recommender systems. In G. M. Olson, & R. Jeffries (Eds.), *Extended abstracts proceedings of the 2006 conference on human factors in computing systems* (pp. 1097–1101). ACM.
- Miller, H. J. (2004). Tobler's first law and spatial analysis. *Annals of the Association of American Geographers*, 94(2), 284–289.
- Neidhardt, J., Seyfang, L., Schuster, R., & Werthner, H. (2015). A picture-based approach to recommender systems. *Journal of Information Technology and Tourism*, 15(1), 49–69.
- Ning, X., Desrosiers, C., & Karypis, G. (2015). A comprehensive survey of neighborhood-based recommendation methods. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 37–76). Springer.
- Ning, X., & Karypis, G. (2011). SLIM: sparse linear methods for top-N recommender systems. In D. J. Cook, J. Pei, W. Wang, O. R. Zaiane, & X. Wu (Eds.), *11th IEEE international conference on data mining* (pp. 497–506). IEEE Computer Society.
- Palumbo, E., Rizzo, G., Troncy, R., & Baralis, E. (2017). Predicting your next stop-over from location-based social network data with recurrent neural networks. In J. Neidhardt, D. R. Fesenmaier, T. Kuflik, & W. Wörndl (Eds.), *CEUR workshop proceedings: Vol. 1906, Proceedings of the 2nd workshop on recommenders in tourism co-located with the 11th ACM conference on recommender systems* (pp. 1–8). CEUR-WS.org.
- Quadrana, M., Cremonesi, P., & Jannach, D. (2018). Sequence-aware recommender systems. *ACM Computer Survey*, 51(4), 66:1–66:36.
- Rafailidis, D., & Crestani, F. (2017). A collaborative ranking model for cross-domain recommendations. In E. Lim, M. Winslett, M. Sanderson, A. W. Fu, J. Sun, J. S. Culpepper, E. Lo, J. C. Ho, D. Donato, R. Agrawal, Y. Zheng, C. Castillo, A. Sun, V. S. Tseng, & C. Li (Eds.), *Proceedings of the 2017 ACM on conference on information and knowledge management* (pp. 2263–2266). ACM.
- Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2009). BPR: Bayesian personalized ranking from implicit feedback. In J. A. Bilmes, & A. Y. Ng (Eds.), *UAI 2009, proceedings of the twenty-fifth conference on uncertainty in artificial intelligence* (pp. 452–461). AUAI Press.
- Sabou, M., Onder, I., Brasoveanu, A. M. P., & Scharl, A. (2016). Towards cross-domain data analytics in tourism: a linked data based approach. *Journal of IT & Tourism*, 16(1), 71–101.
- Sahebi, S., & Brusilovsky, P. (2015). It takes two to tango: An exploration of domain pairs for cross-domain collaborative filtering. In H. Werthner, M. Zanker, J. Golbeck, & G. Semeraro (Eds.), *Proceedings of the 9th ACM conference on recommender systems* (pp. 131–138). ACM.
- Said, A., & Bellogín, A. (2014). Comparative recommender system evaluation: benchmarking recommendation frameworks. In A. Kobsa, M. X. Zhou, M. Ester, & Y. Koren (Eds.), *Eighth ACM conference on recommender systems* (pp. 129–136). ACM.
- Said, A., & Bellogín, A. (2018). Coherence and inconsistencies in rating behavior: estimating the magic barrier of recommender systems. *User Modeling and User-Adapted Interaction*, 28(2), 97–125.
- Sánchez, P., & Bellogín, A. (2018). Challenges on evaluating venue recommendation approaches: Position paper. In J. Neidhardt, W. Wörndl, T. Kuflik, & M. Zanker (Eds.), *CEUR workshop proceedings: Vol. 2222, Proceedings of the workshop on recommenders in tourism, rectour 2018, co-located with the 12th ACM conference on recommender systems* (pp. 37–40). CEUR-WS.org.
- Sánchez, P., & Bellogín, A. (2019). Attribute-based evaluation for recommender systems: incorporating user and item attributes in evaluation metrics. In T. Bogers, A. Said, P. Brusilovsky, & D. Tikk (Eds.), *Proceedings of the 13th ACM conference on recommender systems* (pp. 378–382). ACM.
- Schedl, M., Zamani, H., Chen, C., Deldjoo, Y., & Elahi, M. (2018). Current challenges and visions in music recommender systems research. *IJMIR*, 7(2), 95–116.
- Tang, J., & Wang, K. (2018). Personalized top-n sequential recommendation via convolutional sequence embedding. In Y. Chang, C. Zhai, Y. Liu, & Y. Maarek (Eds.), *Proceedings of the eleventh ACM international conference on web search and data mining* (pp. 565–573). ACM.
- Vargas, S., & Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. In B. Mobasher, R. D. Burke, D. Jannach, & G. Adomavicius (Eds.), *Proceedings of the 2011 ACM conference on recommender systems* (pp. 109–116). ACM.
- Villegas, N. M., Sánchez, C., Díaz-Cely, J., & Tamura, G. (2018). Characterizing context-aware recommender systems: A systematic literature review. *Knowledge-Based Systems*, 140, 173–200.
- Wang, H., Terrovitis, M., & Mamoulis, N. (2013). Location recommendation in location-based social networks using user check-in data. In C. A. Knoblock, M. Schneider, P. Kröger, J. Krumm, & P. Widmayer (Eds.), *21st SIGSPATIAL international conference on advances in geographic information systems* (pp. 364–373). ACM.
- Wattanacharoensil, W., & La-or-nual, D. (2019). A systematic review of cognitive biases in tourist decisions. *Tourism Management*, 75, 353–369.
- Weydemann, L., Sacharidis, D., & Werthner, H. (2019). Defining and measuring fairness in location recommendations. In P. Bouros, T. Dasu, Y. Kanza, M. Renz, & D. Sacharidis (Eds.), *Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based recommendations, geosocial networks and geoadvertising* (pp. 6:1–6:8). ACM.
- Xie, M., Yin, H., Wang, H., Xu, F., Chen, W., & Wang, S. (2016). Learning graph-based POI embedding for location-based recommendation. In S. Mukhopadhyay, C. Zhai, E. Bertino, F. Crestani, J. Mostafa, J. Tang, L. Si, X. Zhou, Y. Chang, Y. Li, & P. Sondhi (Eds.), *Proceedings of the 25th ACM international conference on information and knowledge management* (pp. 15–24). ACM.
- Yang, D., Zhang, D., & Qu, B. (2016). Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM TIST*, 7(3), 30:1–30:23.
- Ye, M., Yin, P., Lee, W., & Lee, D. L. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. In W. Ma, J. Nie, R. A. Baeza-Yates, T. Chua, & W. B. Croft (Eds.), *Proceeding of the 34th international ACM SIGIR conference on research and development in information retrieval* (pp. 325–334). ACM.
- Yu, X., Chu, Y., Jiang, F., Guo, Y., & Gong, D. (2018). SVMS classification based two-side cross domain collaborative filtering by inferring intrinsic user and item features. *Knowledge-Based Systems*, 141, 80–91.

- Yuan, F., Jose, J. M., Guo, G., Chen, L., Yu, H., & Alkhawaldeh, R. S. (2016). Joint geo-spatial preference and pairwise ranking for point-of-interest recommendation. In *28th IEEE international conference on tools with artificial intelligence* (pp. 46–53). IEEE Computer Society.
- Zhang, J., & Chow, C. (2013). iGSLR: personalized geo-social location recommendation: a kernel density estimation approach. In C. A. Knoblock, M. Schneider, P. Kröger, J. Krumm, & P. Widmayer (Eds.), *21st SIGSPATIAL international conference on advances in geographic information systems* (pp. 324–333). ACM.
- Zhang, J., & Chow, C. (2015a). GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In R. A. Baeza-Yates, M. Lalmas, A. Moffat, & B. A. Ribeiro-Neto (Eds.), *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (pp. 443–452). ACM.
- Zhang, J., & Chow, C. (2015b). Spatiotemporal sequential influence modeling for location recommendations: A gravity-based approach. *ACM TIST*, 7(1), 11:1–11:25.
- Zhang, J., Chow, C., & Li, Y. (2014). LORE: exploiting sequential influence for location recommendations. In Y. Huang, M. Schneider, M. Gertz, J. Krumm, & J. Sankaranarayanan (Eds.), *Proceedings of the 22nd ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 103–112). ACM.
- Zhang, C., & Wang, K. (2016). POI recommendation through cross-region collaborative filtering. *Knowledge Information System*, 46(2), 369–387.
- Zhao, Y., Nie, L., Wang, X., & Chua, T. (2014). Personalized recommendations of locally interesting venues to tourists via cross-region community matching. *ACM TIST*, 5(3), 50:1–50:26.
- Zheng, Y. (2015). Methodologies for cross-domain data fusion: An overview. *IEEE Transactions on Big Data*, 1(1), 16–34.
- Zheng, V. W., Zheng, Y., Xie, X., & Yang, Q. (2010). Collaborative location and activity recommendations with GPS history data. In M. Rappa, P. Jones, J. Freire, & S. Chakrabarti (Eds.), *Proceedings of the 19th international conference on world wide web* (pp. 1029–1038). ACM.