

How to Improve the Recommendation's Accuracy in POI Domains?

Luiz Chaves[1], Nícollas Silva[2], Rodrigo Carvalho[1], Adriano C. M. Pereira[2], Diego R. C. Dias[1(✉)], and Leonardo Rocha[1]

[1] Universidade Federal de São João del-Rei, São João del-Rei, Brazil
luizfufsj@gmail.com, {rodrigo,diegodias,lcrocha}@ufsj.edu.br
[2] Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
{ncsilvaa,adrianoc}@dcc.ufmg.br

Abstract. Nowadays, Recommender Systems (RS) have been applied in most of Location-Based Social Networks (LBSNs). In general, these RSs aim to provide the best points-of-interest (POIs) to users, encouraging them to visit new places or explore more of their preferences. Despite the researches advances in this scenario, there is an opportunity for improvements in the recommendation task. The main reason behind it is related to specific characteristics of this scenario, such as the geolocation of users. In general, most users are not interested in POIs located far from their home or work area. In this sense, we address a new research perspective in the POI Recommendation field, proposing a re-ordering method to be applied after any RS and improve the POIs located nearby from the users' geolocation. Our assumption is that POIs located on the sub-areas with more activity of a user are more interesting than POIs from new sub-areas. For this reason, we propose to measure the activity level of users in subareas of a city and use it to re-order the POIs recommended before. We evaluate our proposal considering six traditional RSs and three datasets from Yelp, achieving gains up to 15% of precision.

Keywords: Recommender systems · POI recommendation · Geographic

1 Introduction

Currently, several Web applications have invested in Recommendation Systems (RS) to attract and satisfy more users. On systems like Amazon.com and Netflix, RS's task is to present products and movies that please users [3]. On the other hand, in the case of social networks such as Facebook, RS's task is related to identifying which publications are most relevant to be presented to users [10]. On location-based social networks (i.e., Location-Based Social Network (LBSN)), such as Foursquare, Gowalla, and Yelp, the RS's task is related to identifying users' interests in points of interest (i.e., Points of Interest – POI) such as restaurants, bars, cinemas, theaters, museums, and others [17].

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12249, pp. 557–571, 2020.
https://doi.org/10.1007/978-3-030-58799-4_41

In traditional domain applications, such as e-commerce (products) and entertainment (movies and music), there are already several RS proposals that can achieve their goals of attracting and satisfying users [3,21]. However, in LBSN applications, the recommendation of POIs is still an open challenge due to their peculiar characteristics. Unlike other scenarios, where an item consumption means that the user only watched or bought a particular item, in the POI scenario, it means that the user had to travel to a particular region of the city to visit it. In other words, there is a geographic factor intrinsically related to user activity that must be taken into account to generate the recommendations [16,23]. For this reason, the main work in recommending POIs concerns to obey Tobler's First Law of Geography: *"everything is related to everything else. But near things are more related than distant things"* [18].

The proposals found in the literature consist of traditional RS adaptations to the POI recommendation scenario [24], as well as some new RSs specifically proposed for this purpose [13,23,26]. That is a relatively new research scenario, and none of these approaches have been able to achieve effectively satisfactory results so far [24]. In this sense, instead of proposing a new recommendation model, we propose a new approach to assist existing systems in identifying potentially relevant POI to users. We propose a re-ordering method to be used together with any RS. Our proposal is basically to measure the level of activity of each user in the various subregions that make up a particular area (e.g., city) and, from that level of activity, to weight the recommendation list generated by a recommender according to it. Thus, we ensure that POIs belonging to high-activity subregions receive a higher weight than POIs from low-activity subregions. From this weighting, this list is then reordered and presented to the user.

In the evaluation of our approach, we apply the re-ordering method proposed considering models used in traditional application domains (i.e., MostPopular, User-KNN, and WRMF) and specific models for the POI recommendation problem (i.e., USG, GeoMF, and GeoSoCa). We evaluated all these applications against the Yelp database, widely used in the POI recommendation scenario, by selecting three cities (Las Vegas, Phoenix, and Charlotte). The results show that, although considering a simple metric to measure users' activity level, our strategy was able to improve the recommendation quality in the vast majority of cases analyzed, with some gain higher than 15% accuracy as for the case with the Most-Popular algorithm to the city of Las Vegas. Even applying our solution to POI-specific RS, it was possible to see quality gains, some of which were higher than 8% accuracy, such as the GeoMF to Charlotte City.

The main contribution of this work is a new perspective to address the problem of recommending POIs, introducing a re-ordering method that can be used by any RS. Also, this strategy opens a new line of research, regarding improvements to our proposal (i.e., different strategies to measure user activity level and define subregions) and/or new re-ordering steps considerations. The remainder of this work is organized as follows. In Sect. 2 we present a theoretical framework with the main RSs proposed in POI recommendations. In Sect. 3, we describe our re-ordering proposal, presenting the activity level metric. In Sect. 4, we present

all the details of the experimental project, highlighting the databases used. In Sect. 5, we present the results and discussions inherent in this work. Finally, in Sect. 6, we summarize our conclusions.

2 Background Concepts

In this section, we present recent advances in the recommendation domain, formalizing the problem, and describing the main existing models. Later we describe the specific recommendation context in POIs, where we discuss the main features of their models. Also, we briefly present the main work related to this context.

2.1 Recommender Systems

In general, the recommendation task is to quantify the relevance of an item to a particular user [2,3]. For this reason, many works define the problem as a prediction task where the goal is to estimate the user's rating assigned to an item [4]. On the other hand, the recommendation task is a ranking task, which aims to predict the top-k most relevant items of each user [5]. Both approaches use information passed on by users, such as watched movies, purchased products, logs and/or access cookies, and others, to set a preference profile [11]. Moreover, using the profile to identify and recommend some items that best match users' wishes through some specific approach.

The most used and with the best results among existing approaches is Collaborative Filtering (CF) [20]. CF approaches assume that similar users similarly evaluate items and/or similar items receive similar evaluations from users [3]. Based on this premise, two main classes of methods called memory-based and model-based stand out. The memory-based CF methods are based on the ratings previously provided by the users to recommend items [9]. In this case, user-based or item-based models are highlighted, such as the traditional User-kNN and Item-kNN methods. On the other hand, model-based methods recommend from descriptive models of user preferences, built by strategies derived from machine learning, or even from algebraic [19] models. Matrix Factorization models are considered state of the art.

2.2 POI Recommendation

POIs scenario differs from the others due to their peculiar characteristics Within the recommendation domain. First, many works highlight that the problem of sparsity is even more significant in the POIs domain since the user faces physical obstacles to visit places [16,23]. Besides, other works highlight that the source of information about users differs from others. In this scenario, the primary sources of information to measure user interest are (1) social; (2) temporal; and (3) geographic. Social influence assumes that the opinions of friends present on the user's social network are more relevant than those of non-friends [6,22]. In turn,

the temporal influence refers to the fact that users tend to visit places, for example, restaurants, in an irregular period [14,25]. Above all, geographical influence is the main feature to be considered since it is intrinsic to this recommendation scenario since users and POIs are located in a physical space [24].

While there are several effective recommendation approaches in traditional scenarios, they are not equally effective in all cases. For this reason, we find in the literature approaches that explore the different characteristics of POI recommendations. [7] explored the temporal influence, assuming that users with similar consumption histories, in the same period, share the same interests. Social influence is exploited in [23] and [26]. [23] extracts the influence side by side among the friends of each user. [26] created a cumulative distribution function to estimate social relationships. Finally, the geographical influence is considered the most important, and it presents the best results since the user always takes into account the distance of a place to visit it [12,13,16,23,26]. Among these, we can highlight the works presented by Ye et al. [23] and Lian et al. [13], as those that are considered state of the art. These authors use past information to evaluate the probability of a user visiting a POI or a region [23], given the spread of the influence of POIs in the geographical space [13]. For both cases, the recommendation model considers the probability in the building process (i.e., in the factoring matrix).

Contrasting the works presented in this section, this paper addresses the problem of recommendation from a different perspective. In our case, we propose a re-ordering method for any RS, be it traditional or POI specific, which weights the list of POIs according to the level of user activity in each region. Our solution reorders the list before presenting it to the user. It is an orthogonal proposal to all those presented in this section, that is, it can be used together with all of them, achieving excellent results, as we will see in Sect. 5.

3 Geographical Filtering

Geographic influence is one of the main factors affecting the recommendations of POIs [12,16,23]. In general, users are interested in visiting places near their current location and/or the most visited regions (high level of activity) [24]. For this reason, we propose to filter the recommendations presented by any base model. In general, we intend to quantify the level of user activity in a particular region and use it to drive or penalize some specific recommended POIs. To do this, we first divide a region into sub-areas based on literature suggestions [8]. Then we calculate the level of user activity in each of these sub-regions. Finally, we use this information to process the recommendations made in order to improve them. We call this re-ordering method as Geo-Filtering. We illustrate where our method fits into the POI recommendation process (Fig. 1). We describe further details in the following sections.

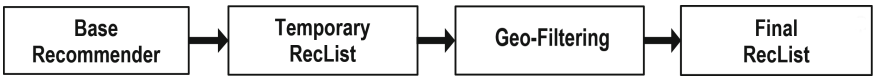


Fig. 1. Steps in the recommendation process

3.1 Recommendation Areas and Sub-areas

Inspired by the recommendation approach adopted by Han & Yamana [8], we initially defined a A recommendation area as a city or district delimited by latitudes and longitudes: initial and final. Then we divide this same A region into fixed-size squared sub-areas with dimensions of 0.5×0.5 km. This way, we created an array of a_{ij} subareas, as we can see in Fig. 2. So, each sub-area a_{ij} covers a portion of A without overlap between the subareas. This 0.5km fixed division is a standard value that we found in the literature [8], and where the experimental tests proved to be most effective.

0.6	0.8	0.7	0.4
0.5	0.9	0.4	0.1
0.2	0.5	0.9	0.2
0.1	0.8	0.7	0.3

Fig. 2. Division into subareas

The idea of dividing A into sub-areas was initially presented in Han & Yamana [8] in order to mitigate the problem of the scarcity of POIs. In this case, the authors proposed a pre-processing that filters the available POIs for the recommendation, leaving only those of the busiest subareas. In turn, in our method, we customized this definition, taking into account the user activity in each region, the location of the POIs, and the users' activity level [14]. Although simple, this division into fixed-size sub-regions has shown excellent results. However, this is one of the stages of our strategy that can still be improved. For example, establishing different dimensions for each sub-region according to the amount of POIs present or even using clustering algorithms by density, such as DBSCAN [1], to establish these sub-regions automatically.

3.2 User's Activity Level

As mentioned in the literature, the main works in POIs recommendation are concerned with obeying the Tobler's First Law. Based on this law, it is possible to derive two fundamental restrictions to achieve user satisfaction [8,26]. Based on them, users prefer:

(R1) visit places near your current location;
(R2) visit places located in regions where they have a high level of activity.

Thus, we propose a metric that we call activity-level (AL), capable of quantifying the user's interest in a sub-area in order to satisfy these two restrictions. Basically, we calculate the user's activity level u in each sub-area $a_i \in A$ by the ratio of the frequency f of u in a over the sum of all check-ins of u (Eq. 1). This information satisfies R2 directly since it measures the user's activity, and indirectly R1 since it uses only the check-ins of the places the user was able to visit.

$$AL_{u,a} = \frac{f_{u,a}}{\sum_{i=1}^{N} f_{u,a_i}} \tag{1}$$

AL represents the proportional frequency of u in the sub-area a. Sub-areas with higher AL values represent those where the user has performed many check-ins during a period comprised of the training data. On the other hand, subareas with AL values equal to or close to 0 represent those where u does not or is infrequent, respectively. It is worth noting that it is possible to expand the way to calculate AL, considering other factors such as social relationships or time characteristics. However, these analyses do not cover the scope of this work.

3.3 Re-ordering Strategy

Our method consists of a new re-ordering approach for POIs recommendation, applied after the calculation of AL values for each user. We assume that a base recommender presents a list of possible N POIs to be visited. Traditionally, this recommender estimates the relevance r_i of each candidate POI according to its implemented methodology and reorders them in a decreasing way of relevance.

From this information, we execute our re-ordering strategy on the temporary list, adding the relevance value of these POIs, original from the base recommender, with the AL ($AL_{u,a}$). Therefore, for each POI in the temporary list, we calculate the $f(u, i)$ utility according to the Eq. 2, which multiplies the relevance notes returned by the base recommender by the AL value of $AL_{u,a}$, where a consists of the sub-area in which the POI is geographically located. At the end of this step, we resort the temporary list according to these new relevant values and select the most relevant k items, where $k \ll N$, as the final list of recommendations *RecList*.

$$f(u, i) = r_i * (1 + AL_{u,a}) \tag{2}$$

In Fig. 3, we present a didactic example of our strategy. In this example, we have the base recommender returning the list of the six most relevant POIs for a user, in this order: $P14$, $P17$, $P1$, $P5$, $P4$ and $P7$. $P1$ is in the third position and has a relevance of 0.7, in sub-region one, which represents the sub-area of the highest user activity with a level equal to 0.7—applying our re-ordering method weighting—we have that the new relevance of $P1$ is given by

$0.7 * (1 + 0.7) = 1.19$. After applying this process to all POIs and reordering the list by the new relevance values, we have POI $P1$ in the first position in the list. We hope that our method will be able to assign more relevance to the POIs that are in subareas where the user has attended more in the past. Also, the Geo-Filtering decreases the final relevance of POIs that are in subareas that have been little frequented by the user, but which could be at the top of the temporary recommendation list provided by the base recommender. This whole process has a small complexity, as it depends on the subregions and POIs frequented by the users and can be calculated in advance.

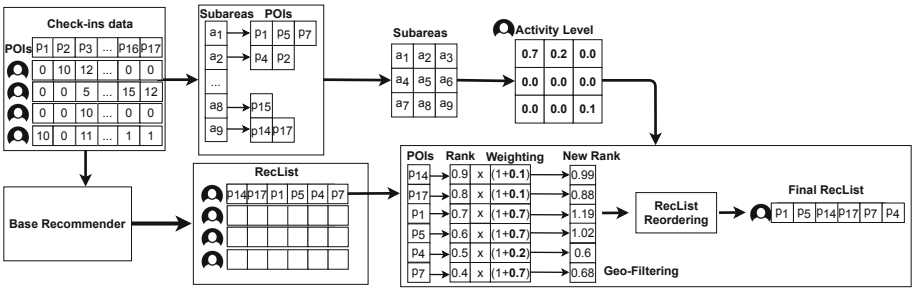


Fig. 3. Re-ordering method application presented in this article

4 Experimental Evaluation

In this section, we describe in detail the assumptions made and the parameters set for our experimental evaluation. First, we highlight the data selection made as proposed by the literature. Next, we describe the basic recommendations that were optimized by our strategy. Finally, we detail the evaluation metrics used to measure the accuracy of the recommendations.

4.1 Datasets

To carry out the experiments, we initially preprocessed the data available in the Yelp Challenge dataset[1]. First, we selected three of the five cities that have the highest number of POIs. These were Charlotte, Las Vegas, and Phoenix. We filtered the dataset information by selecting the POIs that had at least five check-ins and the users that have made at least 20 visits in each city. This process was defined based on the literature to mitigate the problem of user scarcity and maximize the chances of users receiving good recommendations. Thus, after pre-processing, the extracted data are presented in Table 1.

[1] https://www.yelp.com/dataset/challenge.

Table 1. Selected cities from the Yelp dataset.

City	User	POIs	Sparsity
Las Vegas	15.168	11.562	99,752
Phoenix	14.388	36.827	99,849
Charlotte	2.445	7.151	99,239

Then, for each city, we reordered the check-ins according to the creation date, i.e., from the oldest to the most recent. We created a set with the first 70% of the check-ins to train the recommendation algorithms, and another with the remaining 30% to run the tests. We performed this approach to simulate a real scenario of the model, as proposed by Zhang & Chow [26]. It is worth mentioning that we also performed a hierarchical selection to define the category of each POI. Each POI was categorized based on a multilevel category tree. Thus, similar to what was done by Han & Yamana [8], we discarded the categories that are below level 2 in the tree.

4.2 Recommendation Algorithms

To validate our proposal, we have selected the proposed recommendation strategies in two different contexts. First, we selected the classic algorithms implemented in the MyMediaLite library[2]. Then, we selected other proposed approaches for POI recommendation, which are state-of-the-art strategies.

Traditional Recommender Systems:

- **MostPopular:** It recommends a list of the most popular POIs that were not visited yet. The popularity of POIs is the number of visits each POI has.
- **User-kNN:** It recommends the most relevant POIs of the K users most similar to the target user. Users' similarity is calculated based on the cosine similarity, and at least 80 neighbors of the target user are selected.
- **WRMF:** It recommends the best-evaluated POIs based on the association of the user's latent factors with those of the items, modeled by a Matrix Factoring method (e.g., SVD, PCA, and others.).

POIs Recommender Systems:

- **USG:** A technique that recommends POIs taking into account geographical, social, and user preferences. Geographic influence is the possibility of a user visiting a POI given their check-in history and the distances between POIs. Social influence defines the user profile taking into account the preferences of social network friends [23].

[2] http://mymedialite.net/.

- **GeoMF:** It is a matrix factoring recommendation model weighted by geographic information. The geographical factor defines the possibility of a user visiting a region, as well as the influence of each POI on each region [13].
- **GeoSoCa:** It is a recommendation technique that explores the geographic, social, and categorical factors for recommendations. Geographic influence is a kernel estimation method customized by each user's check-ins distribution [26].

4.3 Evaluation Metrics

For the evaluation and comparison of our model with others, we used the traditional metrics of *Precision*@k and *Recall*@k. While *Precision*@k returns the fraction of relevant POIs recovered by the number of recommendations, *Recall*@k measures the fraction of relevant POIs recovered by the amount of POIs consumed in each user's test. We calculated the metrics for each user and returned the average values to evaluate the recommendation effectiveness [26].

5 Results

In this section, we discuss the results obtained by our re-ordering method when applied to selected baselines. We divide our analysis into two cases. First, we evaluate the performance generated by the proposal presented in the traditional recommendation cases, and then, with algorithm specific to POI recommendations.

5.1 Traditional Recommender Systems

First, we selected the three traditional recommendation algorithms called Most-Popular, UserKNN, and WRMF. We applied these algorithms to the Las Vegas, Phoenix, and Charlotte related databases, measuring the Precision and Recall of the recommendations performed. Then, from recommendations generated by each algorithm, we apply our re-ordering process, again measuring the Precision and Recall. We summarize the values found in Table 2.

It is important to emphasize that among these methods, none exploit geographical influence as a decisive factor in recommendations. Therefore, by dividing a region into sub-regions, applying our metrics, geographical filtering can compute information that was not previously explored. With this, we can observe a significant improvement in the results. We found statistically significant gains in the city of Las Vegas when compared to the original list (Table 2). The same happens when we compare the results in the city of Phoenix. Specifically, in the Most-Popular technique, the results are more explicit, as geographic filtering reorders the list in a personalized way to better meet the users' consumption pattern. This way, we were able to recover relevant POIs, which were previously discarded by the traditional recommender. In both cases, Wilcoxon's test confirmed gains for non-normal distributions with a p-value ≤ 0.05.

Table 2. It shows the results achieved by the evaluation metrics with the traditional recommendations and the values achieved by our re-ordering strategy. The symbol ▲ represents statistically significant positive gains, ● represents statistical ties, and ▼ represents statistically significant losses. Wilcoxon's test obtained these gains with a 95% confidence interval

		Precision		Recall		nDCG		MRR	
	Recommender	Top-10	Top-15	Top-10	Top-15	Top-10	Top-15	Top-10	Top-15
Las Vegas	Most popular	0.0153	0.0155	0.0266	0.0395	0.0205	0.0260	0.0450	0.0491
	Geo-filtering	0.0172 ▲	0.0167 ▲	0.0331 ▲	0.0459 ▲	0.0248 ▲	0.0304 ▲	0.0527 ▲	0.0569 ▲
	User-kNN	0.0311	0.0279	0.0583	0.0755	0.0476	0.0552	0.0971	0.1020
	Geo-filt	0.0317 ▲	0.0285 ▲	0.0606 ▲	0.0777 ▲	0.0489 ▲	0.0565 ▲	0.0994 ▲	0.1043 ▲
	WRMF	0.0327	0.0296	0.0592	0.0767	0.0485	0.0567	0.0994	0.1043
	Geo-filtering	0.0332 ▲	0.0301 ▲	0.0613 ▲	0.0794 ▲	0.0500 ▲	0.0582 ▲	0.1027 ▲	0.1077 ▲
Phoenix	Most popular	0.0129	0.0118	0.0089	0.0119	0.0096	0.0115	0.0356	0.0379
	Geo-filtering	0.0131 ▲	0.0123 ▲	0.0090 ▲	0.0127 ▲	0.0098 ▲	0.0120 ▲	0.0361 ▲	0.0387 ▲
	User-kNN	0.0327	0.0294	0.0244	0.0329	0.0276	0.0328	0.0973	0.1022
	Geo-filtering	0.0329 ●	0.0298 ▲	0.0247 ▲	0.0336 ▲	0.0279 ●	0.0332 ●	0.0979 ●	0.1030 ●
	WRMF	0.0276	0.0255	0.0204	0.0278	0.0226	0.0272	0.0797	0.0845
	Geo-filtering	0.0281 ▲	0.0259 ▲	0.0208 ▲	0.0284 ▲	0.0231 ▲	0.0277 ▲	0.0812 ▲	0.0858 ▲
Charlotte	Most popular	0.0367	0.0338	0.0280	0.0386	0.0290	0.0352	0.0971	0.1024
	Geo-filtering	0.0375 ▲	0.0344 ▲	0.0285 ●	0.0400 ▲	0.0297 ●	0.0361 ●	0.0978 ●	0.1033 ●
	User-kNN	0.0481	0.0441	0.0399	0.0544	0.0416	0.0500	0.1317	0.1378
	Geo-filtering	0.0478 ▼	0.0443 ●	0.03963 ●	0.0547 ●	0.0416 ●	0.0503 ●	0.1325 ●	0.1390 ●
	WRMF	0.0492	0.0444	0.0393	0.0525	0.0424	0.0503	0.1375	0.1433
	Geo-filtering	0.0499 ●	0.0450 ●	0.0404 ●	0.0539 ▲	0.0432 ●	0.0512 ●	0.1392 ●	0.1450 ●

The gains earned by our re-ordering stage in the cities of Las Vegas and Phoenix were not observed for the city of Charlotte. In the vast majority of combinations, our re-ordering stage shows a statistical tie to the original algorithm results. The first observation is that Charlotte is the city with the smallest number of users and an equally small amount of POIs per sub-area. Also, the number of POIs per user is low. The little consumption information from the users considerably affects the quality of the traditional recommender systems. As a consequence, the list of recommended POIs tends to contain items that are not very relevant for users. Our strategy is purely dependent on the recommendation list and the area of operation of each user. Thus, reordering a list of POIs that are mostly no longer user-relevant items has little effect on the final quality. In any case, this result shows that there is still room for other re-ordering proposals to be used in conjunction with the strategy presented here.

5.2 POIs Recommendation

In our second analysis, we selected the three baselines regarding the recommendation of POIs, which are considered state of the art in this field. Specifically, we selected USG [23], GeoMF [13] and GeoSoCa [26]. Again, we retrieved the generated recommendation list and applied our geographic filtering proposal. All the results obtained can be seen in Table 3.

Table 3. It shows the results achieved by the evaluation metrics with the specific recommendations for POIs and the values achieved by our re-ordering strategy. The symbol ▲ represents significant positive gains, ● represents statistical ties, and ▼ represents significant losses. Wilcoxon's test obtained these gains with a 95% confidence interval

		Precision		Recall		nDCG		MRR	
	Recommender	Top-10	Top-15	Top-10	Top-15	Top-10	Top-15	Top-10	Top-15
Las Vegas	USG	0.0276	0.0257	0.0545	0.0721	0.0442	0.0520	0.0889	0.0939
	Geo-filtering	0.0290 ▲	0.0266 ▲	0.0583 ▲	0.0759 ▲	0.0466 ▲	0.0544 ▲	0.0926 ▲	0.0975 ▲
	GeoMF	0.0284	0.0255	0.0546	0.0713	0.0437	0.0510	0.0896	0.0942
	Geo-filtering	0.0291 ▲	0.0261 ▲	0.0551 ●	0.0732 ▲	0.0446 ▲	0.0522 ▲	0.0923 ▲	0.0971 ▲
	GeoSoCa	0.0206	0.0180	0.0339	0.0439	0.0282	0.0326	0.0660	0.0691
	Geo-filtering	0.0201 ▼	0.0177 ●	0.0322 ▼	0.0430 ●	0.0271 ▼	0.0318 ▼	0.0628 ▼	0.0661 ▼
Phoenix	USG	0.0228	0.0212	0.0176	0.0241	0.0198	0.0238	0.0701	0.0744
	Geo-filtering	0.0231 ▲	0.0215 ▲	0.0179 ▲	0.0246 ▲	0.0202 ▲	0.0243 ▲	0.0714 ▲	0.0758 ▲
	GeoMF	0.0297	0.0278	0.0235	0,0330	0.0256	0.0312	0.0874	0.0928
	Geo-filtering	0.0302 ▲	0.0281 ▲	0.0242 ▲	0.0336 ▲	0.0263 ●	0.0318 ●	0.0893 ●	0.0946 ●
	GeoSoCa	0.0171	0.0154	0.0112	0.0151	0.0131	0.0156	0.0516	0.0543
	Geo-filtering	0.0171 ●	0.0155 ●	0.0111 ●	0.0153 ●	0.0131 ●	0.0157 ●	0.0512 ●	0.0539 ●
Charlotte	USG	0.0420	0.0387	0.0334	0.0457	0.0338	0.0409	0.1088	0.1143
	Geo-filtering	0.0432 ▲	0.0394 ▲	0.0347 ▲	0.0475 ▲	0.0350 ▲	0.0423 ▲	0.1111 ●	0.1172 ▲
	GeoMF	0.0349	0.0325	0.0282	0.0391	0.0300	0.0364	0.0992	0.1050
	Geo-filtering	0.0357 ●	0.0334 ▲	0.0299 ▲	0.0409 ▲	0.0313 ●	0.0377 ▲	0.1028 ▲	0.1084 ▲
	GeoSoCa	0.0230	0.0218	0.0163	0.0235	0.0183	0.0226	0.0630	0.0674
	Geo-filtering	0.0237 ●	0.0215 ●	0.0167 ●	0.0232 ●	0.0189 ●	0.0227 ▲	0.0636 ▲	0.0676 ▲

In this case, our model achieved excellent results for the recommendation techniques USG and GeoMF in the three selected cities. It is because our proposal can assist the POIs specific recommendation algorithms, which deal with the geographical factor. Basically, in these cases, the activity level defined by the proposal of this work aggregates the geographic information in the recommendations, taking into account the activity density of each user in sub-regions. It is possible to pronounce the most difference in the re-ordering of the USG in Las Vegas and Charlotte. Again, we obtained statistically significant gains, with the Wilcoxon test (p-value ≤ 0.05). These results reinforce the idea that our proposal achieves the expected goals and contributions, especially on the GeoMF model, one of the main ones in the literature [15].

By comparing our proposal with GeoSoCa [26], the re-ordering performance decreased, both in Las Vegas and Phoenix. It is because GeoSoCa adds two more attributes to its recommendations: social and categorical influences. Furthermore, according to Liu et al. [15], the social factor explored by GeoSoCa is the one with the best performance compared to other algorithms (e.g., the USG). Therefore, our proposal deals only with the geographic characteristics of each user individually, thus justifying that difference found. However, even being inferior, the results obtained were very close to those found by the baseline, and was not considered statistically significant losses.

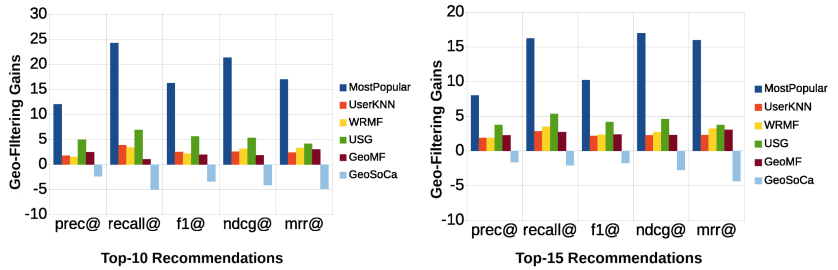
In the city of Charlotte, we observed an impressive result. We found gains concerning the strategy of GeoSoCa. It is because Charlotte has less social

correlation data, where each user, on average, has only one friend. Therefore, the aggregation of information that the social factor brings to GeoSoCa is not relevant. Thus, the application of our re-ordering stage could enrich the geographic information bringing improvements to the recommendation of GeoSoCa.

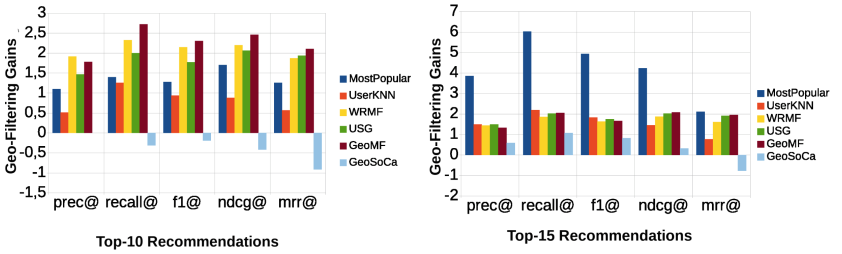
5.3 Discussions

In this section, we present an analysis of the gains obtained by each algorithm, in each city, when using the Geo-Filtering re-ordering proposal presented in this work. These results are all summarized in Fig. 4. The first interesting observation is the gains obtained by the MostPopular algorithm (i.e., gains of up to 15% on accuracy for the city of Las Vegas). It is a non-personalized strategy, widely used in Pure Cold Star Problem scenarios, where the recommender selects the most consumed items in the training base for the users. In this case, Geo-Filtering ends up introducing customized information, which makes the quality of this algorithm improve significantly. In smaller proportions, Geo-Filtering is also able to improve the quality of other traditional classifiers (i.e., UserKNN and WRMF). We have observed that the most significant quality gains occur in cities where the average consumption of users is high. The consumption information is very relevant for these algorithms because it is from this information that one can model the profile of users and from this information to make recommendations of items, in this case, POIs, that have similar characteristics to the past consumption of users. However, this process of consumption modeling does not take into account the restrictions imposed by Tobler's geography law, and it is precisely this gap that our re-ordering strategy fills, adding to these algorithms geographic information of users' consumption.

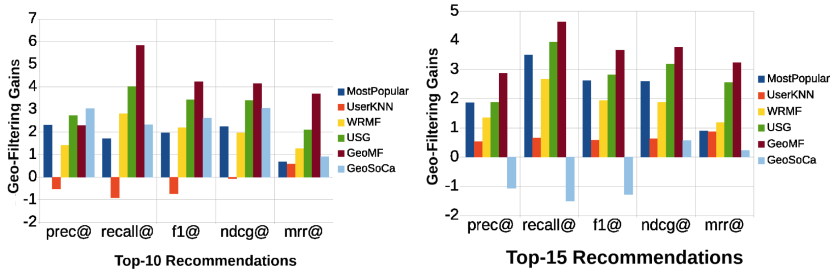
As we have seen throughout this section, we also considered state-of-the-art algorithms in the area of POIs recommendation (i.e., USG, GeoMF, and GeoSoCa). These algorithms also incorporate the geographic information from the POIs into the generation of the models (i.e., the factoring matrix). Thus, our objective was to evaluate if Geo-Filtering was still able to aggregate some new information that was not explored by these algorithms. Observing the gains obtained (Fig. 4), we concluded that yes. For these cases, we hypothesize that Geo-Filtering not only aggregates geographic information, but it is also able to correlate this information to the users' consumption. The GeoSoCa algorithm is a clear exception. As previously mentioned, it uses other factors like information, especially the social factor, which, according to the authors, is the one that most contributes to the proper functioning of the algorithm. However, when we look at the gains that Geo-Filtering can obtain for GeoSoCa for the city of Charlotte (i.e., gains higher than 8% accuracy), where there is almost no social information, we realize that the geographical factor is still little explored by GeoSoCa. Thus, we believe that the most significant contribution of this work is precisely in opening a new perspective of research in the area of POIs recommendation, introducing new re-ordering stages.



(a) Las Vegas



(b) Phoenix



(c) Charlotte

Fig. 4. Summarization of the gains obtained by the geo-filtering application in each algorithm/city combination

6 Conclusions and Future Works

In this work, we propose a re-ordering model that weighs up and filters a set of recommendations generated by a base recommender system in order to further satisfy user preferences. To do it, we build a metric capable of quantifying the level of user activity in sub-regions. Our premise is that by satisfying the two constraints determined by Tobler's geography law, we further favor user satisfaction. In general, we observe that the inclusion of Geo-filtering, as re-ordering, allows, and effectively, to increase the accuracy of recommendations. This new re-ordering stage adds features not yet explored by other baselines. Dividing an area into smaller sub-areas and calculating for each user their level of activity,

makes it possible to extract the most active sub-regions from users. The results show statistically significant gains in most cases, especially in large cities, such as Phoenix and Las Vegas.

Our proposal is still limited to baselines that exploit other features than geography, such as the GeoSoca. Our goal, as future work, is to include other characteristics as well, such as the social influence in the calculation of the user's activity level in order to extract more information not evaluated by the RSs. We also intend to include other ways to obtain the subareas, dividing them in a varied way or including new subareas within those already divided.

Acknowledgments. This work was partially funded by the Brazilian National Institute of Science and Technology for the Web - INWeb, MASWeb, CAPES, CNPq, Finep, Fapesp and Fapemig.

References

1. Andrade, G., Ramos, G., Madeira, D., Sachetto, R., Ferreira, R., Rocha, L.: G-DBSCAN: a GPU accelerated algorithm for density-based clustering. Procedia Comput. Sci. -ICCS **18**, 369–378 (2013)
2. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005)
3. Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A.: Recommender systems survey. Knowl.-Based Syst. **46**, 109–132 (2013)
4. Chikhaoui, B., Chiazzaro, M., Wang, S.: An improved hybrid recommender system by combining predictions. In: 2011 IEEE Workshops of International Conference on Advanced Information Networking and Applications (WAINA), pp. 644–649 (2011)
5. Cremonesi, P., Donatacci, A., Garzotto, F., Turrin, R.: Decision-making in recommender systems: the role of user's goals and bounded resources. In: Decisions@ RecSys, pp. 1–7 (2012)
6. Fernández-Tobías, I., Tomeo, P., Cantador, I., Di Noia, T., Di Sciascio, E.: Accuracy and diversity in cross-domain recommendations for cold-start users with positive-only feedback. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 119–122. ACM (2016)
7. Gao, H., Tang, J., Hu, X., Liu, H.: Exploring temporal effects for location recommendation on location-based social networks. In: Proceedings of the 7th ACM Conference on Recommender Systems, pp. 93–100. ACM (2013)
8. Han, J., Yamana, H.: Geographical diversification in POI recommendation: toward improved coverage on interested areas. In: Proceedings of the Eleventh ACM Conference on Recommender Systems, pp. 224–228. ACM (2017)
9. Hu, L., Sun, A., Liu, Y.: Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 345–354. ACM (2014)
10. Kumar, S.S., Ramachandran, T., Panboli, S.: Product recommendations over Facebook: the roles of influencing factors to induce online shopping. Asian Soc. Sci. **11**(2), 202 (2015)

11. Lacerda, A., Ziviani, N.: Building user profiles to improve user experience in recommender systems. In: Proceedings of the 6th ACM WSDM, pp. 759–764 (2013)
12. Li, X., Cong, G., Li, X.L., Pham, T.A.N., Krishnaswamy, S.: Rank-GeoFM: a ranking based geographical factorization method for point of interest recommendation. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 433–442. ACM (2015)
13. Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., Rui, Y.: GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 831–840. ACM (2014)
14. Liu, B., Fu, Y., Yao, Z., Xiong, H.: Learning geographical preferences for point-of-interest recommendation. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1043–1051. ACM (2013)
15. Liu, Y., Pham, T.A.N., Cong, G., Yuan, Q.: An experimental evaluation of point-of-interest recommendation in location-based social networks. Proc. VLDB Endow. **10**(10), 1010–1021 (2017)
16. Liu, Y., Wei, W., Sun, A., Miao, C.: Exploiting geographical neighborhood characteristics for location recommendation. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, pp. 739–748. ACM (2014)
17. Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G.: Recommender system application developments: a survey. Decis. Support Syst. **74**, 12–32 (2015)
18. Miller, H.J.: Tobler's first law and spatial analysis. Ann. Assoc. Am. Geogr. **94**(2), 284–289 (2004).https://doi.org/10.1111/j.1467-8306.2004.09402005.x
19. Mourão, F.H.d.J.: A hybrid recommendation method that combines forgotten items and non-content attributes (2014)
20. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72079-9_9
21. Silva, N., Carvalho, D., Pereira, A.C., Mourão, F., Rocha, L.: The pure cold-start problem: a deep study about how to conquer first-time users in recommendations domains. Inf. Syst. **80**, 1–12 (2019)
22. Ye, M., Yin, P., Lee, W.C.: Location recommendation for location-based social networks. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 458–461. ACM (2010)
23. Ye, M., Yin, P., Lee, W.C., Lee, D.L.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: Proceedings of the 34th International ACM SIGIR, pp. 325–334. ACM (2011)
24. Yu, Y., Chen, X.: A survey of point-of-interest recommendation in location-based social networks. In: Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
25. Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M.: Time-aware point-of-interest recommendation. In: Proceedings of the 36th International ACM SIGIR, pp. 363–372. ACM (2013)
26. Zhang, J.D., Chow, C.Y.: GeoSoCa: exploiting geographical, social and categorical correlations for point-of-interest recommendations. In: Proceedings of the 38th International ACM SIGIR, pp. 443–452. ACM (2015)