

Selecting and Weighting Users in Collaborative Filtering-based POI Recommendation

Carlos Ríos, Silvia Schiaffino and Daniela Godoy

ISISTAN Research Institute (CONICET-UNCPBA)
Campus Universitario, Paraje Arroyo Seco
CP 7000, Tandil, Bs. As., Argentina
{carlos.rios,silvia.schiaffino,daniela.godoy}@isistan.unicen.edu.ar

Abstract: Location-based recommender systems (LBRs) provide a technological solution for helping users to cope with the vast amount of information coming from geo-localization services. Most online social networks capture the geographic location of users and their points-of-interests (POIs). Location-based social networks (LBSNs), like Foursquare, leverage technologies such as GPS, Web 2.0 and smartphones allow users to share their locations (check-ins), search for POIs, look for discounts, comment about specific places, connect with friends and find the ones who are near a specific location. LBRs play an important role in social networks nowadays as they generate suggestions based on techniques such as collaborative filtering (CF). In this traditional recommendation approach, prediction about a user preferences are based on the opinions of like-minded people. Users that can provide valuable information for prediction need to be first selected from the complete network and, then, their opinions weighted according to their expected contribution. In this paper, we propose and analyze a number of strategies for selecting neighbors within the CF framework leveraging on information contained in the users' social network, common visits, visiting area and POIs categories as influential factors. Experimental evaluation with data from Foursquare social network shed some light on the impact of different mechanisms on user weighting for prediction.

Keywords: Location-based social networks; recommender systems; user-based collaborative filtering.

1 Introduction

Recent technological advances in the development of wireless communications, the great explosion of cell phone use, and the easiness to acquire the geographical location of people, have allowed the creation of social services whose main feature is the geographical location of users. In this new era, users can benefit from obtaining a pervasive and ubiquitous access to location-based services from anywhere

through mobile devices. Foursquare¹ is the most popular location-based social media service [11], allowing users to easily share their geographical location as well as contents related to that location in an online way. The user location is a new dimension in social networks that created new opportunities and challenges for traditional recommendation systems. Recommender systems are an alternative to deal with the problem of information overload that users face while seeking information about items of interest in vast amounts of knowledge. Traditional methods such as collaborative filtering (CF), content-based recommendation (CB) and hybrid methods [29] process information obtained from the ratings provided by users and the characteristics of the items involved to generate a list of recommendations. However, in social networks there is additional information that recommendation methods should take into account, such as users' behavior and relations of friendship between them [38].

Regarding location-based social networks (LBSNs), geo-localized data is a physical dimension that traditional social networks do not possess. In this context, location-based recommender systems (LBRs) have emerged [5] as a means to exploit geographical properties as an auxiliary source for recommending friends [25, 31], places [23], activities [39, 7] and events [27, 15]. The heterogeneity of the data produced by location-based social networks creates the need for new approaches in recommendation systems, using different data sources and methodologies for enhancing recommendations. The Collaborative Filtering (CF) approach, for example, relates users to items through ratings or opinions, so that it can be straightforwardly applied to the construction of LBRs. However, the traditional CF approach lacks the geo-localization dimension.

In this work, we propose different strategies for including the additional dimensions available in LBSNs in the context of user-based collaborative filtering for recommending locations. In a LBSN, there are relationships of various types, such as the User-User relationship, showing the friendship between two users or the coincidence in places visited by these users; the User-Place relationship showing that a user visited a given place; the Place-Place relationship, which shows distance relationships between places or categorical membership. Also, in addition to these relationships, users generate content-based relations by providing comments or tips after visiting a place.

User-based approaches recommend items (e.g. places) based on an aggregation of the preferences of similar users or neighbors, i.e. users with similar tastes. As user-based CF trusts neighbors as information sources, the quality of recommendations is a direct consequence of the selected neighborhood and the importance given to each neighbor for prediction. In a previous work [30], we have studied different strategies for selecting neighbors considering the information contained in the LBSNs. We have concluded that selecting as neighbors those users that have visited the same places as the target user we can obtain lower errors in the preference estimation process. To weight neighbors, some works have considered factors such as trust or geographical or social influence [3]. In this work, our main hypothesis is that location-based social networks provide rich information that can enable us to

¹ <http://es.foursquare.com/>

give neighbors representative weights and, hence, improve the estimation of preferences during the recommendation process. We propose and evaluate nine strategies using real data from a Foursquare dataset. From the empirical evaluation, we found that some of the proposed alternatives outperformed the traditional approach, giving developers some hints about which aspects to consider when making recommendations.

The rest of the article is organized as follows. Section 2 presents an overview of the CF approach for POI recommendation, describing the framework in which the proposed strategies for selecting and weighting neighbors fits in. Section 3 describes the experimental results we carried out to evaluate the different strategies and our findings. Then, Section 4 analyzes some related works. Finally, in Section 5 we present our conclusions and outline some future works.

2 CF-based POI recommendation

In a traditional CF scenario, there are m users $U = u_1, u_2, \dots, u_m$, and a list of n items $I = i_1, i_2, \dots, i_n$, that can be recommended to users. Each user has expressed her opinion about a set of items $I_{u_i} \subseteq I$, generally in an explicit way with a rating or value in a given numerical scale. This information is stored in a user-item matrix M of size $m \times n$, such that the value of each cell in M represents the preference score (rating) given by user i to item j . Memory-based CF approaches make predictions based on the user-item matrix in two ways, based on users or based on items [1]. Given an active user who requires a prediction for an item without rating, CF algorithms measure the similarities between the active user and other users (user-based approach), or between the item and the remaining items (item-based approach). Therefore, a rating is predicted by an aggregation of the ratings that the item received from similar users in the first case, or ratings given by the active user to similar items in the second case.

The classic user-based CF model is then defined as in Equation 1 [28]:

$$\tilde{r}(u, i) = \bar{r}(u) + C_o \sum_{v \in N_k(u, i)} sim(u, v)(r(v, i) - \bar{r}(v)) \quad (1)$$

where $r(v, i)$ is the rating given by user v to item i , \tilde{r} is the rating prediction (different from the observed rating r), $N_k(u, i)$ is the set of k most similar users to u and $sim(u, v)$ is the function that determines the similarity between users u and v . C_o is a normalizing factor. The preference of user u for an item i is predicted according to the average rating $\bar{r}(u)$, the sum of deviations of the ratings given by the neighbors v to item i and the average ratings $\bar{r}(v)$, weighting by the similarity with neighbors.

User-based approaches assume that not all users are equally useful in the prediction for a given user, thus two main problems emerge: (1) selecting neighbors for a user to generate recommendations; (2) how to use properly the information provided by those neighbors in the generation of recommendations. Usually, the selection of

neighbors is based on their similarity to the active user, while a common practice is to define a maximum number of users to narrow the neighborhood. Once the neighborhood is defined, the contribution of each neighbor to the prediction is weighted based on their distance from the active user. For example, a widely used alternative is a linear combination of the ratings weighted by the similarity of the neighbors. However, there are other factors that may be valuable for selecting neighbors. For example, in the case of this work the users' history of visits can be considered relevant beyond the ratings similarity.

To properly separate the two problems, the selection of neighbors on the one hand and the weighting of their opinions on the other hand, [6] proposes a modification of the classic formula. This new formula considered an allocation score function (scoring) depending on the active user u , a neighbor v and an item i , or some combination thereof. This function gives a higher value when the triplet of user-neighbor-item is more valuable or expected to work better in predicting a rating according to the available information. Eq. 1 is then generalized as Equation 2:

$$\tilde{r}(u, i) = \bar{r}(u) + C_o \sum_{v \in g(u, i; k; s)} f(s(u, i, v), \text{sim}(u, v)) * (r(v, i) - \bar{r}(v)) \quad (2)$$

where g is the function that selects neighbors and f is an aggregation function that combines the outcomes of the scoring function s and $\text{sim}(u, v)$ the similarity between users.

The selection of neighbors involves the determination of the similarity of users to the target user, by making a comparison with all the users in the database. So any user that is similar to the target user may contribute to the preference estimation. The function g (selection of neighbors) may be influenced by relations present in a LBSN. Thus, restricting with some criteria the potential neighborhood of a target user by exploiting the information generated in LBSN, we can reduce the number of comparisons and, at the same time, improve the preference estimation.

Regarding the scoring function s , we can use different information available in the LBSN to determine which neighbors are better predictors and hence, improve the preference estimation. For example, it can be assumed that users visiting the same places are more useful for prediction than those visiting a different set of places. Likewise, we can suspect that users establishing a friendship relationship are more valuable as a source for estimating preferences. In this context, we present different approaches for the scoring function once neighbors are selected and evaluate them empirically.

2.1 Selection of Neighbors

In a location-based social network users can be related by common visited places. In this paper, we used a strategy for selecting neighbors based on graph of common visited places. In a previous work [30], this strategy outperformed other strategies considered for selecting neighbors.

Particularly, it was compared with a strategy that selects neighbors from the user social network (friendship relationships), and two strategies based on the geographical location of users, one choosing users from the same state and other based on the intersection of visiting areas.

The strategy based on common visited places assumes that users that have visited the same places as the target user are the most valuable for asking opinions. Thus, a graph is generated starting from the preference matrix where nodes represent users and an edge between two nodes implies that both users have visited at least one place in common. The relations in the graph are weighted by the number of coincides in the same places normalized considering the total number of places visited by the user. Equation 3 shows how to obtain the value of the relationship $R(u_1, u_2)$ between two users u_1 and u_2 . p_{u1} and p_{u2} are the places visited by users u_1 and u_2 respectively. In order to keep only strong relationships between users, we discarded those edges with a weight below the average.

$$R(u_1, u_2) = \frac{|p_{u1} \cap p_{u2}|}{|p_{u1} \cup p_{u2}|} \quad (3)$$

Figure 1 shows an example of the graph formation. Once the graph has been created, it is traversed for selecting the best possible neighbors. The graph traversal can be performed up to several depth levels. In other words, the first level will be formed only by users that have visited the same places as the target user. After the first level, users can be selected if they are related to users in the first level. As the graph is explored further, the number of selected neighbors grows. Consequently, not only the computational cost of computing prediction is higher, but also noisy users can be incorporated. In the experimental evaluation performed, the extent to which the graph should be explored is analyzed.

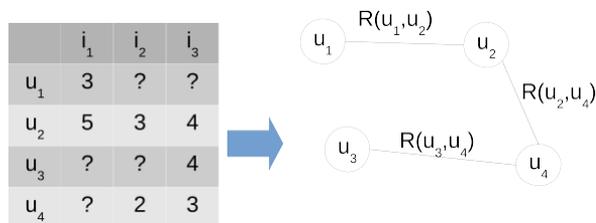


Figure 1
Example of a graph of common visiting places

In the graph exemplified by Figure 1, there is not edge between users u_1 and u_3 as they have no common places. The stronger relationship is between users u_2 and u_4 as $R(u_2, u_4) = 0.40$, whereas the remaining ones are $R(u_1, u_2) = 0.25$ and $R(u_3, u_4) = 0.33$.

2.2 Strategies for Weighting Neighbors

In this article, we propose nine scoring or neighbor weighting strategies considering different pieces of information contained in a LBSN. These strategies are described below in different groups, those based on a discretization of the preference matrix (Subsection 2.2.1), those based on the proximity of the area visited by users (Subsection 2.2.2), those leveraging the categories of POIs (Subsection 2.2.3) and those using the friendship graph and the graph of common visited places (Subsection 2.2.4).

2.2.1 Scoring based on preference matrix discretization

This strategy proposes to derive two new matrices from the original preference matrix by discretizing user preferences. These matrices are called “positive” and “negative” and they represent a coincidence both in the places visited and in the positive or negative user preference towards places. The rationale behind these strategies is that the commonalities on positive and negative preferences makes users more valuable for estimating the preferences of new places.

A first strategy is based on the positive matrix exclusively and a second one is based on both positive and negative matrices. In the last case, we use Jaccard similarity to combine both matrices and obtain a metric that represents the correspondence in negative and positive preference. Jaccard similarity is shown in Equation 4. Figure 2 shows an example of user preference discretization. In this example, a threshold value of 3 in a five-point scale is considered to classify preferences into positive or negative ones. Equations 5 and 6 represent these scoring strategies.

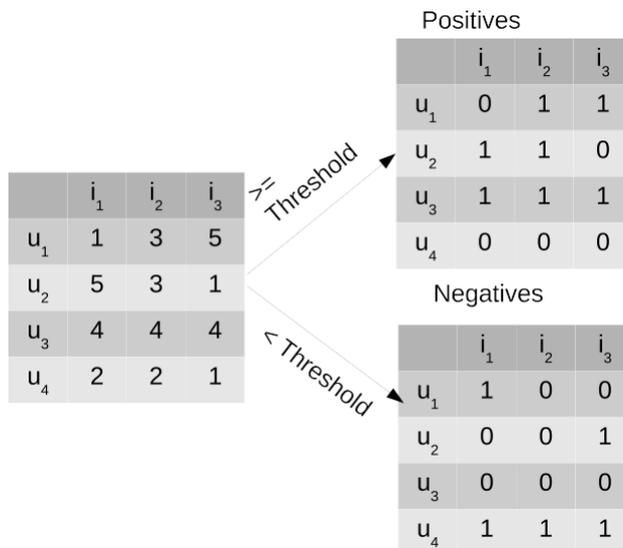


Figure 2
Matrix binarization process

$$Jaccard(u, v) = \frac{|u \cap v|}{|u \cup v|} \quad (4)$$

$$s1(u, v, i) = s(u, v) = Jaccard(u, v)_{PositiveMatrix} \quad (5)$$

$$s2(u, v, i) = s(u, v) = mean(Jaccard_{PositiveMatrix}, Jaccard_{NegativeMatrix}) \quad (6)$$

2.2.2 Scoring based on users' area proximity

In this strategy, the area visited by users is used for weighting higher those users that are moving in the same region. For each user, we first obtain the geographical area that covers the places most visited by the user. Then, the scoring is computed by assessing the distance between the areas for the different users. Equation 7 shows the scoring function for this strategy, where $dist(areaU, areaV)$ is the geodesic distance between the centers of each area.

$$s3(u, v, i) = s(u, v) = 1/dist(areaU, areaV) \quad (7)$$

According to this equation, the closer the areas of both users the higher the scoring. Thus, the users visiting areas less distant to each other are considered more relevant for obtaining opinions.

2.2.3 Scoring based on POIs categories

POIs can be categorized depending on the application domain according to a hierarchy. For example, the dataset we used contains a hierarchy of three levels, where the first level contains general categories such as "Arts & Entertainment", "Food", and "Nightlife Spot". The second level contains more specific categories, such as "Arcade" and "Art Gallery" for "Arts & Entertainment"; "South American Restaurant" or "Taco Place" for "Food"; and "Strip Club" and "Whisky Bar" for "Nightlife Spot". Then, the third level contains lower-level categories such as "Indie Movie Theater", "Paella Restaurant", "Rock Club", etc.

Leveraging POIs categories we define strategies for weighting neighbors based on two approaches. The first approach derives from the original preference matrix another matrix that shows places grouped by the lowest category in the hierarchy. This approach enables to find more intersections between users than the original matrix, since we are not considering specific places (a certain restaurant) but the category to which they belong to (e.g. paella restaurant).

The second approach, proposes a specialization of the original preference matrix according to the categories in the top level of the hierarchy. Thus, if the hierarchy has N categories in the first level, we obtain N preference matrices, one for each category. Then, to compute the scoring of neighbors, we need information about the user, the neighbor and the POI category. This second strategy aims at giving more importance to those users that visit places in the same category. Equations 8 and 9 show the scoring functions proposed for these two strategies, where $Jaccard(u, v)$ is Jaccard similarity and $sim(u, v)$ is any other similarity metric.

Figure 3 shows an example of the proposed strategies. In Figure 3(a) the original matrix is transformed into a matrix in which all of the users have coincidences on items belonging to two of the lowest categories SC_{11} and SC_{21} , that could be for example “Indie Movie Theater” in one branch of the hierarchy and “Paella Restaurant” in other branch. Figure 3(b) shows instead the translation of the original matrix to one matrix by category. One matrix collects the ratings on a main category Cat_1 , that could be for example “Arts & Entertainment”, and other matrix gathers the rating of another category Cat_2 , for instance “Food”. Thus, both matrices can be used separately depending of the category of the target item to generate predictions.

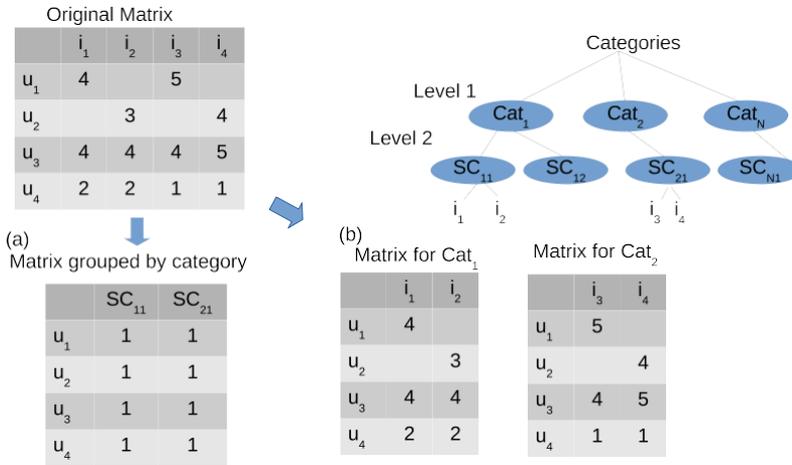


Figure 3
Extraction of preference matrices based on categories

$$s4(u, v, i) = s(u, v) = Jaccard(u, v)_{MatrixGroupedByCategory} \quad (8)$$

$$s5(u, v, i) = sim(u, v)_{CategoryMatrix-i} \quad (9)$$

2.2.4 Scoring based on a graph relationships

This strategy proposes the utilization of structural similarity metrics between two nodes in a network. In our domain, we can use the social graph representing friendship relationships between users. Also, we can build a graph in which nodes are users and the relationships among them represent common visited places by the users as the one used in the neighbor selection strategy. In the social graph, two users are similar if they have friends in common, and in the other graph two users are similar if they are related to users that visited the same places. Equations 10 and 11 show how to compute the scoring functions for these strategies, where $N(u)$ and $N(v)$ are the neighbors of u and v in the graph, respectively.

$$s6(u, v, i) = s(u, v) = \sigma_{jaccard(u,v)} = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|} \quad (10)$$

$$s7(u, v, i) = s(u, v) = \sigma_{coseno(u,v)} = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)| |N(v)|}} \quad (11)$$

Both Jaccard and Cosine local similarity can be also applied to calculate the similarity of two users in the graph of common visited places. We denote these strategies $s8(u, v, i)$ and $s9(u, v, i)$, respectively.

2.2.5 Summary of strategies

To sum up, the strategies we propose and analyze in this article are summarized below:

- $s1(u, v, i)$ generates positive and negative preference matrices based on the discretization of the preference values. This scoring function uses only the positive sentiments, favoring users having coincidences in the positive matrix.
- $s2(u, v, i)$ generates positive and negative preference matrices based on the discretization of the preference values. This scoring function uses both sentiments, favoring users having coincidences in both positive and negative matrices.
- $s3(u, v, i)$ considers areas of POIs visited more frequently by users, favoring users having bigger area intersection, i.e. the closer a person is the more important its opinions.
- $s4(u, v, i)$ considers the hierarchically organized POIs categories and the number of coincidences in the lowest-level categories of POIs visited by the two users.
- $s5(u, v, i)$ considers the hierarchically organized POIs categories as well as the coincides of users in the highest-level category the target item belongs to.

- $s6(u, v, i)$ calculates the Jaccard local similarity index over the social graph (friendship relationships).
- $s7(u, v, i)$ calculates the Cosine local similarity index over the social graph (friendship relationships).
- $s8(u, v, i)$ calculates the Jaccard local similarity index over the graph representing common visited places.
- $s9(u, v, i)$ calculates the Cosine local similarity index over the graph representing common visited places.

3 Experimental results

In this section, the different experiments we carried out to evaluate the performance of the different strategies are proposed. We compared the Mean-Absolute Error (Equation 12) for the different neighbor selection alternatives against the classical user-based collaborative filtering approach. First, in Section 3.1 we describe the characteristics of the dataset used. Then, in Section 3.2 we present the results obtained and analyze them.

$$MAE = \frac{1}{|T|} \sum_{(u,i) \in T} | \bar{r}_{ui} - r_{ui} | \quad (12)$$

The baseline chose for comparing the results of the defined strategies as well as the initial selection of neighbors is the traditional user-based CF approach in which the k most similar users are selected by calculating the cosine similarity of preferences between the target user and all users in the system and the selected users are weighted according to the same similarity. For calculating the MAE rates the preference matrix was divided into a 70% for training and 30% for testing, i.e. the error estimation is done over the last group considering the profiles in the former one.

3.1 Dataset description

For the experiments the dataset from [4] was used, containing data collected from one of the most widely used LBSN, *Foursquare*. The dataset has the following information: *Places*, information about the places visited; *Users*, data of the users using the system; *Tips*, information about check-ins made by users; *Friendship*, information on the social relationship between users, and *Categories*; information of the categories of *Foursquare* places. In the dataset there are users from all over the world, but for our experiments only users belonging to the state of New York were considered, as they are greater in quantity. Out of the 47,220 users in the dataset, the 27,000 users from New York were used.

As explained in Section 2, collaborative filtering relies on a rating matrix M where each cell represents the preference score (rating) given by a user i to item j . For

extracting such a matrix from the dataset, preferences need to be inferred from the user actions in the dataset since there is not explicit rating given to places. User tips, however, constitute implicit information about the user's preferences and they can be processed in order to interpret the meaning of the text and the sentiment associated to it to obtain a preference score.

To this end, in this work we use an automatic sentiment analysis tool to obtain a numerical value denoting the users' opinion about the places they visited expresses in the tips left. Sentiment analysis or opinion mining is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes [21]. This discipline has a lot of activity in recent years since the proliferation of social media and their applications [2]. Given the practical implications of this task in the automatic analysis of the content generated in social media, such as reviews, forum discussions or posts, multiple tools have emerged for extracting sentiment from input texts.

User tips were processed to obtain the preference matrix. For each user, all tips belonging to the same place were concatenated into a single text for tuning the sentiment analysis tool. TextBlob² tool was used to obtain the opinion or sentiment corresponding to the text extracted from tips. This tool was used as it offers a simple API for diving into common natural language processing (NLP) tasks. TextBlob is a python library that among other functionalities, contains a sentiment analysis function to extract the sentiment of a given text. The sentiment is expressed in a $[-1, 1]$ range, where -1 means that the sentiment or opinion is negative, and 1 means that the sentiment or opinion is positive. Values between the extremes, indicate different degrees of positiveness or negativeness. Finally, the sentiments were mapped to a five-point scale for completing matrix M . Table 1 shows how the values for sentiments are discretized.

Table 1
Discretization of sentiment analysis values extracted from tips

Sentiment value	Preference value
$[-1,-0.6)$	1
$[-0.6,-0.2)$	2
$[-0.2,0.2)$	3
$[0.2,0.6)$	4
$[0.6,1]$	5

The process of transforming the preference matrix obtained as described into a graph of common visiting places, illustrated in Figure 1, results in a network containing 37,679 nodes and 440,436 edges. Experiments were run for evaluating the selection of neighbors over this graph and, then, the impact of the different scoring functions for the estimation of preferences.

² <https://textblob.readthedocs.io/en/dev/index.html>

3.2 Results

In the graph of common visiting places built as described in the previous sections, the selection of neighbors is first addressed. For these experiments, different neighborhood sizes were considered, from 5 users up to 300 users. This upper limit was chosen as results tend to stabilize as the neighborhood size grows. Figure 4 shows the results achieved by selecting users at different depth levels. Also, the results are summarized for different neighborhood sizes. In the figure, it is possible to observe that independently of the depth explored, the selection of neighbors based on the graph outperformed the baseline. The best results were obtained using relationships of level 1, i.e. considering the opinion of users that visited the same places as the target user. After level 1, the selection of neighbors does not achieve better results in terms of prediction. Figure 5 shows the average number of users extracted from the graph for determining the neighborhood of each target user. From level 2 to 5 the number of users involved in preference estimation during prediction is higher, but this does not cause a reduction of estimation errors. In average, in the first level an average of 27 users are considered, whereas at level 5 an average of 31661 users are extracted from the graph for making a prediction.

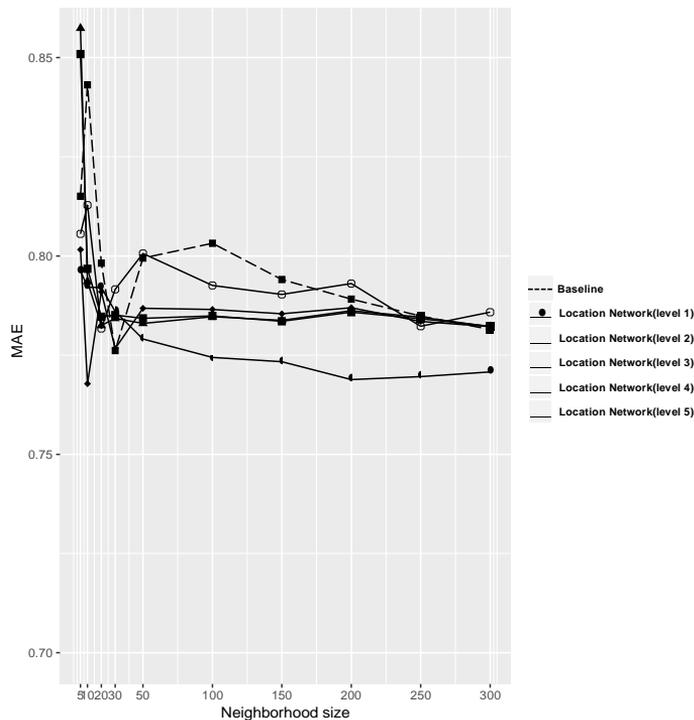


Figure 4
Results of selecting neighbors in the graph of common visited places

Scoring strategies are compared after selecting neighbors from the graph of common visited locations. Figure 6 shows the results obtained using as neighbor selection the

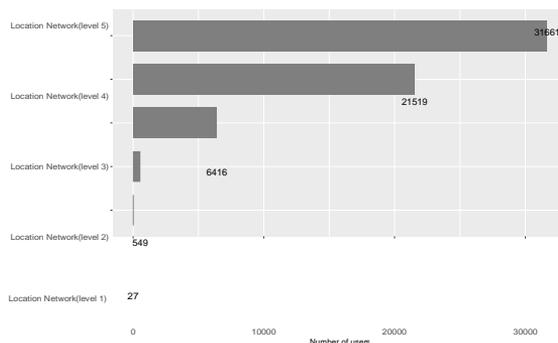


Figure 5

Average number of users compared when selecting the neighborhood

network of common visited places at different levels (from 1 to 5) and the different scoring strategies proposed in this article, compared against the baseline obtained by selecting the k most similar neighbors using cosine similarity.

In general, we can observe that the deeper the network is explored the higher the MAE, as noticed in the previous experimentation. This could be due to the fact that considering more neighbors in the preference estimation introduces some noise in the calculus. We can also observe that in level 1 of the network, i.e. direct relationships, all the scoring strategies have less MAE than the baseline, being strategies $s5$ and $s2$ the ones that produce less error differences. In the remaining levels, most of the strategies are worse than the baseline. Strategy $s5$ considers the item category, thus the weight that of each neighbor considered is more specific regarding the item. Strategy $s2$ takes into account the positive and negative coincidences in user preferences, thus a user that has more coincidences both in places and in opinions with the target user has a higher weight.

Regarding level 2 in the network, we can observe that strategies $s5$ and $s2$ achieved less error values than the baseline, but the other scoring strategies increased their error being even higher than the baseline in some configurations. Also, in level 2 the best overall results are achieved for strategy $s2$, implying that including more users (an average of 549 in such level), but appropriately weighted, can lead to a better prediction. In level 3, the tendency is similar than in level 2. In levels 4 and 5 in the network, $s2$ still has a less MAE than the baseline, but $s5$ increased its value being similar to the baseline and other strategies.

4 Related works

POIs recommendation plays an important role in LBSNs as it helps users discover and explore new attractive locations taking advantage of the community-contributed data, such as friendship links between users, check-ins on points-of-interest, comments, geographical information and categories of POIs, all of them reflecting user preferences. LBSNs, therefore, open new possibilities and challenges for recommender systems [5].

Several POI recommendation systems have been proposed in the literature stem-

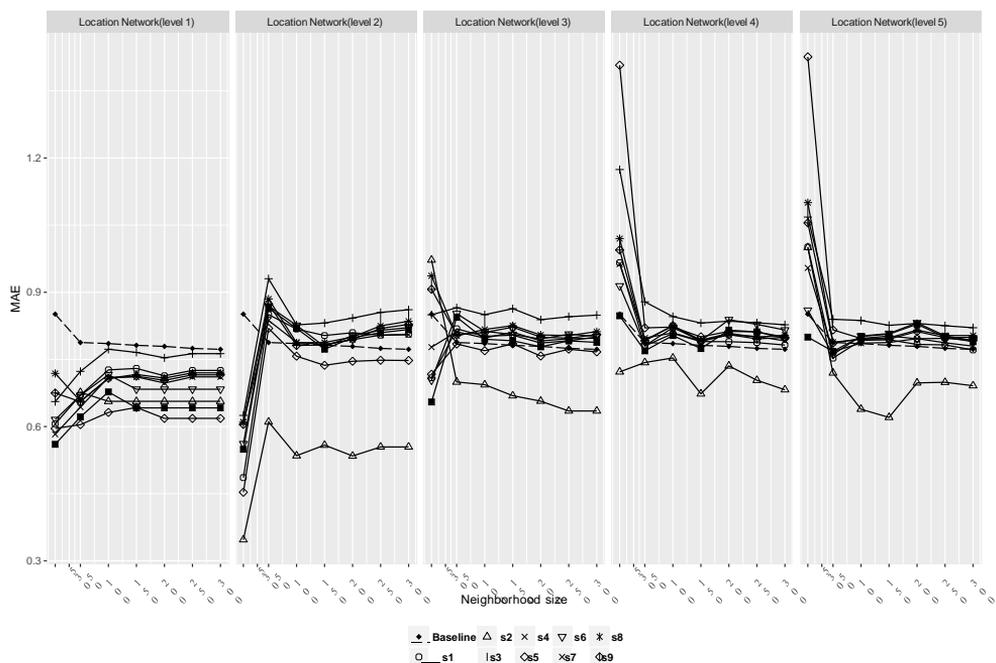


Figure 6
Results of the different neighbor weighting strategies

ming from traditional techniques from the area of recommender systems such as content-based, collaborative filtering and hybrid approaches. LCARS [36], a location-content-aware recommender system, exploits the content information about a user preferred spatial items to produce recommendations in other cities. In [7], the authors propose an approach for detecting the current user context, inferring possible leisure activities and recommend appropriate content on the site (shops, parks, movies). [34] explore text descriptions, photos, user check-in patterns, and venue context for defining a location semantic similarity for venue recommendation. The Sequential Personalized spatial items REcommender system (SPORE) [33] fuses the sequential influence of visited spatial items and the personal interests of individual users using a novel latent variable topic-region model. In [19, 18] the problem of cold-start (i.e., recommending locations to new users) is addressed with a hybrid content-aware collaborative filtering approach that exploits the rich semantics (e.g., tweets) that users often shared on social networks. [40] propose a cross-region collaborative filtering (CRCF), to address both long-term content preferences and short-term location preferences of users.

Specifically, in the context of collaborative filtering, there are works that translated user check-ins into a user-item matrix, where each row corresponds to a user visiting history and each column is a POI. LARS [16], a location-aware recommender system, deals with three types of location-based ratings: spatial ratings for non-

spatial items, non-spatial ratings for spatial items, and spatial ratings for spatial items. Berjani et al. [8] applied a Regularized Matrix Factorization (RMF) technique for CF-based personalized recommendation of potentially interesting spots. [20] incorporates into the factorization model a spatial clustering phenomenon observed in human mobility behavior on LBSNs. Within the CF framework, [13] use highly available GPS trajectories to enhance visitors with context-aware POI recommendations and [42] extract the user travel experience in the target region to reduce the range of candidate POIs. [37] introduce the temporal behavior of users into a time-aware POI recommendation and [41] propose an opinion-based POI recommendation framework taking advantage of the user opinions on POIs expressed as text-based tips.

Regarding users' weighting, various authors have proposed methods for integrating information about user's connections into recommender systems in the form of "trust-based recommenders". Massa and Avesani [3, 24] proposed a method for using users' statements of their trust in other users' opinions to weight user ratings by estimated trust rather than similarity in user-user CF when producing predictions and recommendations and built a ski mountaineering site around it. Golbeck [10] used a similar integration method, based on a different trust estimation algorithm, for movie recommendation. In [14] the authors present a variation of k-NN, the trusted k-nearest recommenders (or k-NR) algorithm, which allows users to learn who and how much to trust one another by evaluating the utility of the rating information they receive. This way users are no longer weighted according to how close to like-minded are, but according to the quality of the information that they exchange with each other. Rating information is weighted according to trust, a value that reflects a history of interactions rather than a history of similar ratings. There is continued work, however, on various methods for estimating and propagating trust through social networks [3, 24, 26, 43]. Guy et al. [12] also found that users tend to find recommendations of web sites, discussion forums, and other social software more interesting when they were recommended from the user's social connections rather than users with similar preference histories.

Particularly in the context of LBSN, the influence that some users may have on location recommendations for other users on account of social or spatial relationships has been addressed in different works. In [35] the authors consider the social and spacial influence of users under the framework of user-based CF, introducing this influence into a model-based method (a Bayesian CF algorithm). Trust and distrust relationships are used to identify friends for recommendation rather than considering all users. The authors found that geographical influence has a significant impact on the accuracy of POI recommendations, whereas the social friends contribute little to the accuracy. Liu et al. [22] incorporated instance and region level of geographical neighborhood characteristics into the learning of latent features of users and locations. Gao et al. [9] propose the concept of geo-social correlations of users' check-in activities, which considers both social networks and geographical distance to model four types of social correlations (i.e., local friends, distant friends, local non-friends and distant non-friends). These correlations are used for solving the "cold start" location prediction problem. In [32] a local context is defined, modeling the correlation between users and their friends, and a global context, denoting the

reputation of users in the social network that is employed to weight the importance of user ratings. Li *et al.* [17] define three types of friends in LBSNs, social friends, location friends, and neighboring friends. Then, a two-step framework leverage the information of friends to improve POI recommendation in the context of a matrix factorization model.

In contrast to the described works, this paper aims to assess the impact of the different strategies proposed in the scoring of neighbors in user-based CF-based POI recommendation. In a previous work [30], strategies for selecting neighbors were studied, showing the advantage of building a graph of common visited places. The scoring strategies evaluated considered the different elements available in LBSNs, such as places categories and social information.

5 Conclusions

In this article, we have proposed different strategies for neighbor scoring in the context of collaborative filtering for the recommendation places of interest (POIs) in LBSNs. We have combined these strategies with a neighbor selection approach that is based on a graph of common visited places. This strategy outperformed others in the experimental evaluation reported in a previous work [30]. We carried out different experiments to evaluate the strategies proposed against a traditional user based collaborative filtering approach. Some of the strategies obtained very low errors values in the estimation of user preferences such as the strategy that considers the hierarchically organized POIs categories as well as the coincides of users in the highest-level category the target item belong to, and the strategy that is based on coincidences in positive and negative sentiment matrices based on the discretization of the preference values. These findings might be useful for recommender system developers in the context of LBSNs, since they can consider our results to prioritize different aspects or dimensions present in these networks to make recommendations. In future works, we plan to extend the experimentation on larger-scale datasets as well as incorporate other possible elements available in different location-based social networks.

Acknowledgements

The work described in this article has been partially funded by ANPCyT, Argentina, through project PICT 2011-0366.

References

- [1] G. Adomavicius and A. Tuzhilin: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(6), pp. 734–749.

- [2] O. Appel, F. Chiclana, and J. Carter: Main concepts, state of the art and future research questions in sentiment analysis, *Acta Polytechnica Hungarica*, 2015, 12(3), pp. 87–108.
- [3] P. Avesani, P. Massa, and R. Tiella: A trust-enhanced recommender system application: Moleskiing, In *Proceedings of the 2005 ACM Symposium on Applied Computing (SAC 2005)*, Santa Fe, New Mexico, USA, 2005, pp. 1589–1593.
- [4] J. Bao, Y. Zheng, and M. Mokbel: Location-based and preference-aware recommendation using sparse geo-social networking data, In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12)*, Redondo Beach, USA, 2012, pp. 199–208.
- [5] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel: Recommendations in location-based social networks: A survey, *Geoinformatica*, 2015, 19(3), pp. 525–565.
- [6] A. Bellogín, P. Castells, and I. Cantador: Neighbor selection and weighting in user-based collaborative filtering: A performance prediction approach, *ACM Transactions on the Web*, 2014, 8(2), pp. 12:1–12:30.
- [7] V. Bellotti, B. Begole, E. H. Chi, N. Ducheneaut, J. Fang, E. Isaacs, T. King, M. W. Newman, K. Partridge, B. Price, P. Rasmussen, M. Roberts, D. J. Schiano, and A. Walendowski: Activity-based serendipitous recommendations with the Magitti mobile leisure guide, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Pages (CHI '08)*, Florence, Italy, 2008, pp. 1157–1166.
- [8] B. Berjani and T. Strufe: A recommendation system for spots in location-based online social networks, In *Proceedings of the 4th Workshop on Social Network Systems (SNS '11)*, Salzburg, Austria, 2011, pp. 4:1–4:6.
- [9] H. Gao, J. Tang, and H. Liu: gSCorr: Modeling geo-social correlations for new check-ins on location-based social networks, In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM '12)*, Maui, Hawaii, USA, 2012, pp. 1582–1586.
- [10] J. Golbeck: Generating predictive movie recommendations from trust in social networks, In *Proceedings International Conference on Trust Management, Lecture Notes in Computer Science*, Springer, 2006, Vol. 3986, pp. 93–104.
- [11] J. Golbeck: Introduction to social media investigation: A hands-on approach, chapter Foursquare, Syngress Publishing, 2015, pp. 101–113.
- [12] I. Guy, N. Zwerdling, D. Carmel, I. Ronen, E. Uziel, S. Yogev, and S. Ofek-Koifman: Personalized recommendation of social software items based on social relations, In *Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys '09)*, New York, NY, USA, 2009, pp. 53–60.
- [13] H. Huang and G. Gartner: Using context-aware collaborative filtering for POI recommendations in mobile guides, In *Proceedings of the 8th International*

- Symposium on Location-Based Services (LBS 2011), Vienna, Austria, 2011, pp. 131–147.
- [14] N. Lathia, S. Hailes, and L. Capra: Trust-based collaborative filtering, In Joint iTrust and PST Conferences on Privacy, Trust Management and Security (IFIPTM), 2008, Vol. 263, pp. 119–134.
- [15] R. Lee, S. Wakamiya, and K. Sumiya: Discovery of unusual regional social activities using geo-tagged microblogs, *World Wide Web*, 2011, 14(4), pp. 321–349.
- [16] J.J. Levandoski, M. Sarwat, A. Eldawy, and M. F. Mokbel: LARS: A location-aware recommender system, In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering (ICDE '12), Washington, DC, USA, 2012, pp. 450–461.
- [17] H. Li, Y. Ge, R. Hong, and H. Zhu: Point-of-interest recommendations: Learning potential check-ins from friends, In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16), San Francisco, CA, USA, 2016, pp. 975–984.
- [18] D. Lian, Y. Ge, N. Jing Yuan, X. Xie, and H. Xiong: Sparse bayesian content-aware collaborative filtering for implicit feedback, In Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI-16), New York, NY, USA, 2016, pp. 1732–1738.
- [19] D. Lian, Y. Ge, F. Zhang, N. Jing Yuan, X. Xie, T. Zhou, and Y. Rui: Content-aware collaborative filtering for location recommendation based on human mobility data, In Proceedings of the IEEE International Conference on Data Mining (ICDM2015), 2015, pp. 261–270.
- [20] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui: 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 (KDD '14), New York, NY, USA, 2014, pp. 831–840.
- [21] B. Liu and L. Zhang: Mining Text Data, chapter A Survey of Opinion Mining and Sentiment Analysis, Springer, Boston, MA, USA, 2012, pp. 415–463.
- [22] Y. Liu, W. Wei, A. Sun, and C. Miao: Exploiting geographical neighborhood characteristics for location recommendation, In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM '14), Shanghai, China, 2014, pp. 739–748.
- [23] X. Long and J. Joshi: A HITS-based POI recommendation algorithm for location-based social networks, In IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013), Niagara, Canada, 2013, pp. 642–647.

- [24] P. Massa and P. Avesani: Trust-aware collaborative filtering for recommender systems, *Lecture Notes in Computer Science*, Springer 2004, Vol. 3290, pp. 275–301.
- [25] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos: Friendlink: Link prediction in social networks via bounded local path traversal, In *International Conference on Computational Aspects of Social Networks (CASoN 2011)*, Salamanca, Spain, 2011, pp. 66–71.
- [26] A. Popescul, L. H. Ungar, D. M. Pennock, and S. Lawrence: Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments, In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence (UAI'01)*, San Francisco, CA, USA, 2001, pp. 437–444.
- [27] D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft: Recommending social events from mobile phone location data, In *IEEE 10th International Conference on Data Mining (ICDM 2010)*, Sydney, Australia, 2010, pp. 971–976.
- [28] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl: Grouplens: An open architecture for collaborative filtering of netnews, In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (CSCW '94)*, 1994, pp. 175–186.
- [29] F. Ricci, L. Rokach, and B. Shapira: *Recommender Systems Handbook*, chapter Introduction to recommender systems handbook, Springer, 2011, pp. 1–35.
- [30] C. Rios, S. Schiaffino, and D. Godoy: On the impact of neighborhood selection strategies for recommender systems in LBSNs, In *Proceedings of the 15th Mexican International Conference on Artificial Intelligence (MICAI 2016)*, Cancun, Mexico, 2016.
- [31] S. Scellato, A. Noulas, and C. Mascolo: Exploiting place features in link prediction on location-based social networks, In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*, San Diego, CA, USA, 2011, pp. 1046–1054.
- [32] J. Tang, X. Hu, H. Gao, and H. Liu: Exploiting local and global social context for recommendation. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI '13)*, Beijing, China, 2013, pp. 2712–2718.
- [33] W. Wang, H. Yin, S. Sadiq, L. Chen, M. Xie, and X. Zhou: SPORE: A sequential personalized spatial item recommender system, In *Proceedings of the IEEE 32nd International Conference on Data Engineering (ICDE 2016)*, Helsinki, Finland, 2016, pp. 954–965.
- [34] X. Wang, Y-L. Zhao, L. Nie, Y. Gao, W. Nie, Z-J. Zha, and T-S. Chua: Semantic-based location recommendation with multimodal venue semantics, *IEEE Transactions on Multimedia*, 2015, 17(3), pp. 409–419.
- [35] M. Ye, P. Yin, W-C. Lee, and D-L. Lee: Exploiting geographical influence for collaborative point-of-interest recommendation, In *Proceedings of the 34th*

- International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '11), Beijing, China, 2011, pp. 325–334.
- [36] H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen: LCARS: A location-content-aware recommender system, In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13), Chicago, IL, USA, 2013, pp. 221–229.
- [37] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. Magnenat-Thalmann: Time-aware point-of-interest recommendation, In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13), 2013, pp. 363–372.
- [38] R. Zafarani, M. Ali Abbasi, and H. Liu: *Social Media Mining: An Introduction*, Cambridge University Press, 2014.
- [39] A. Zanda, E. Menasalvas, and S. Eibe: A social network activity recommender system for ubiquitous devices, In Proceedings of the 11th International Conference on Intelligent Systems Design and Applications (ISDA 2011), Cordoba, Spain, 2011, pp. 493–497.
- [40] C. Zhang and K. Wang: POI recommendation through cross-region collaborative filtering, *Knowledge and Information Systems*, 2016, 46(2), pp. 369–387.
- [41] J-D. Zhang, C-Y. Chow, and Y. Zheng: ORec: An opinion-based point-of-interest recommendation framework, In Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM 2015), Melbourne, Australia, 2015, pp. 1641–1650.
- [42] E. Zhou, J. Huang, and X. Xu: A point-of-interest recommendation method based on user check-in behaviors in online social networks, In Proceedings of the 4th International Conference on Computational Social Networks (CSoNet 2015), Lecture Notes in Computer Science, Springer 2015, Vol. 9197, pp. 160–171.
- [43] C-N. Ziegler and G. Lausen: Propagation models for trust and distrust in social networks, *Information Systems Frontiers*, 2005, 7(4), pp. 337–358.