An Effective POI Recommendation in various Cold-start Scenarios

Pramit Mazumdar
National Institute of Technology Rourkela
Odisha, India
pramitmazumdar@gmail.com

Bidyut Kr. Patra
National Institute of Technology Rourkela
Odisha, India
patrabk@nitrkl.ac.in

Korra Sathya Babu
National Institute of Technology Rourkela
Odisha, India
ksathyababu@nitrkl.ac.in

ABSTRACT

The advances in GPS enabled mobile devices has led people to share their real-time experiences ubiquitously through various online platforms. Users' experiences on various services (products), interesting places (famously known as point of interests (POI)), events, movies, etc. are being widely collected these days by various enterprise houses. It helps the community to browse through these online contents before selecting a product or a POI. Recommender system is proven to be a successful tool which can automatically provide an effective list of items to an active user based on her preferences by filtering through the large item space. However, a recommender system often fails to learn the user preferences for a new user who has no historical data (widely known as the cold start problem). This work focuses on developing a POI recommender system for handling various cold start scenarios such as, 'new user' and 'new city'. In this regard, a Feature and Region based POI Recommender System (FRRS) has been devised which can effectively provide a list of top-K POIs to an active user in cold start scenarios. The proposed system FRRS has two modules, modelling and recommendation. First, the user preferences and features of POIs are learnt from various online contents such as ratings and reviews. Finally, the recommendations are obtained by combining the learnt user preferences with the interests of influential users and the proximity of POIs from the active location for recommending a list of top-K POIs. Experiments are performed on the real-world Yelp dataset. We compare the performance of our approach with three existing works and a baseline approach for recommending POIs. The obtained results show that our proposed approach outperforms the existing works.

1. INTRODUCTION

The growth of location aware technologies has helped people to share information and experiences related to their place of visit. This has been exceedingly utilised by popular social networks such as Facebook and Twitter, which offer geotagging along with each post. In addition to that, there exist location based social networks such as Foursquare and Gowalla which crowd-sources the user generated contents of various point of interests (POI). Many other social networking cites such as Yelp, TripAdvisor, etc. provide online platforms (both browser and mobile based) where business oriented data are collected. Users share their experiences through reviews, ratings and tips on various POIs. Similarly, individual business owners also create their own online profile (facebook page, twitter handle or a website) with detailed information on latest discounts, customer reviews, ratings, etc. Therefore, availability of such detailed and updated information on the web has led people to exploit the online contents before visiting a POI. However, reviewing the online contents manually is a tedious task.

The recommender system help users by providing a personalized list of items from a large item space. The collaborative filtering has long been used for item recommendations. However, it has been observed that the traditional collaborative filtering technique alone is not suitable for POI recommendation [1]. A user who starts using a recommender system for the first time, has no historical data. A collaborative filtering technique of inferring user preferences only from the historical data is not feasible for a new user. Therefore, it fails to provide appropriate recommendations to a new user - widely known as the cold start scenario. Many researcher focused on cold start scenarios in POI recommendations [2, 3, 4]. It has been observed that, exploiting the contents at various POIs along with collaborative filtering is by far a superior approach for POI recommendation. In this work, a feature extraction technique has been employed to model user preferences. Subsequently, the positive ratings and reviews provided to the POIs are utilised to find their representative features. Two cold start scenarios are addressed in this work. First, the user cold start problem termed as the 'new user' problem. Second, the 'new city' problem which is a variant of the standard item cold start problem, where the city is 'new' to the active user. In addition to the cold start problems, the proposed work also provides distinct recommendation strategies for frequent and occasional travellers of a city. Throughout the rest of the paper, the terms 'active user', 'active location' and 'active region' will refer to the user for whom POIs are being recommended, the user's current location and the region within which the user's current location resides, respectively. The contributions of the proposed Feature and Region based POI Recommender System (FRRS) are summarized as below.

1. A feature-centric approach is used to predict the unobserved ratings. The user preferences on specific features associated with a POI are learned using the standard matrix factorization technique.

2. The geographical space is clustered into regions and a set of trusted users for each region are identified. The 'new user' problem is addressed by utilizing the preferences of these users of a region.

3. The 'new city' problem for a user is addressed by identifying the socially influential users for an active user. Collaborative filtering on ratings of the influential users help to estimate the rating of a POI at a region 'new' to the active user.

4. The frequent and occasional travellers to a city are classified into separate category of active user. Distinct recommendation strategies are employed for each of them.

5. Experiments were performed on Yelp dataset to evaluate the proposed recommender system. The results are compared with a baseline approach [5] and three existing approaches [2, 3, 4]. The obtained results show that our proposed recommender system outperforms the existing approaches.

The rest of the paper is organized as follows. In Section 2, we provide details of the existing research works on POI recommendation. A detailed description of the existing approaches with which we compare our results are also produced. Motivation of our current research and the limitations of the existing approaches are presented in Section 3. Certain well known techniques used in the proposed approach are reproduced in Section 4. The proposed framework of FRRS is depicted in Section 5. First, the proposed methodology for extracting preferences of users and features of POIs are mentioned. Subsequently, the technique used for rating various POIs are described. Section 6 reports the results obtained after comparing the existing approaches with the proposed approach. Finally, we conclude our work in Section 7.

2. RELATED WORK

In this section, we provide a detailed description of the existing recommender systems.

The collaborative filtering has been the traditional method for recommendations [6, 7]. The user based collaborative filtering technique for POI recommendation identifies the similar users and subsequently selects the highly rated POIs by them for recommendations [6, 8, 9]. For a new user there exists very less historical data to identify similar users and/or common POIs between them. This is commonly known as the cold start problem. Moreover, limited coverage of the recommended POIs is another issue with the traditional collaborative filtering technique for POI recommendations. Limited coverage refers to a problem where a POI is never considered for recommendations if it is not rated by atleast a neighbor. The matrix factorization technique has also been widely used for recommendations [10, 11, 12]. In a POI recommender system, the content or feature of the respective item/POI plays a very important role. The traditional task of finding 'few good items' is mostly driven by the correlation between content of POIs and preferences of a user. A content based recommender system matches the preferences of a user with the features of POIs [13, 14, 15]. This approach has been found to be providing better results than the traditional collaborative filtering in cold start scenarios. Recent works in [2, 3, 4] combine both the content and collaborative filtering techniques for POI recommendations.

Ye *et al.* [2] provide POI recommendations for the location based social networks, such as Foursquare and Whrrl. The user based collaborative filtering technique is used to extract the user preferences from their activities. Similarly, the friend based collaborative filtering is utilized to identify the social influence on a user. They perform a spatial analysis to state that the geographic proximity of POIs effect a users' activity. Subsequently for each user, the pair-wise distance between the checked-in POIs are computed. It is observed that a larger percentage of check-ins are performed within short distances. This concludes that the proximal locations have high chance of visiting. Hence, the user check-in activities tend to be geographically clustered around a region. Finally a fusion between the three parameters user preference, socio influence and the geographic influence is performed to assign a score for each POI with respect to the active location based social network users. Top-K POIs having high scores are finally selected for recommendation. The recommender system proposed in [2] is henceforth referred as USG in this paper.

Yin *et al.* [3] models personal interest of users and their local preferences for venue and event recommendations. It addresses three problems, data sparsity problem, the 'new city' problem and recommending top-K events around the current location of an active user. The proposed Location Content Aware Recommender System (LCARS) comprises of two parts. First, it models the user preferences in offline mode. A probabilistic model is exploited to learn the interests of each user and the local preferences of each region. Subsequently, the spatial items or POIs within each city are treated as words and feature extraction is performed on them. The distribution of POIs checked-in by each user in a city is learnt from the probabilistic model. Similarly, the model also learns the preferential POIs in a city from the obtained review data. For the 'new city' problem, the popular POIs of a city helps in recommendation. By combining the content preferences and location preferences, a score with respect to a user is provided for each POI in a city. Second, the online recommendation combines the knowledge obtained from user interests and local preferences of each city, to provide top-K recommendations. The POIs having high score for a user are selected for recommendation.

Recent work by Zhang *et al.* [4] propose a cross-region POI recommendation technique. Here the POIs are recommended on basis of two important factors, the past historical data and the active location. To achieve this, the recommendation technique first identifies both the long term and short preferences of each user. A user's liking on certain features (content of a POI) is identified as its long term preference. In this regard, the historical data is explored, and is believed to be consistent (thus, long term). At this point they generate a user-feature matrix. The set of features are identified from the existing feature based ratings provided by various users at POIs. A Matrix-Factorization (MF) technique is employed to obtain the unobserved ratings on a POI feature by a user. The average rating provided by a user on a feature depicts its long term preference on it. For each user the top-m POIs possessing the highly rated features are thus selected for the content-aware recommendations. In addition to it, the location preference of a user is termed as the short term preference. It relates to the proximal POIs with respect to the active location. The observed rating is then adjusted by a distance term which penalizes a POI on basis of its distance from the user's current location. The top-m content aware POIs are arranged on basis of their location preferences. The final list of top-n ($n < m$) POIs are recommended from the set of m POIs to the user. The Cross-Region Collaborative Filtering (CRCF) technique proposed in [4], thus models both the user's content preferences (long term) and location-aware preferences (short term) for recommending effective POIs. It mostly focuses on the 'new city' and data sparsity problem for POI recommendations.

The existing approaches fail to consider the effect of social influences. Moreover, the recommendations are not targeted towards specific category of users. In the next section, the limitations of the existing approaches are described along with the description of each category of user.

3. MOTIVATION AND PROBLEM STATEMENT

A user's choice of visiting a POI depends on three vital factors. First, the preferences or likings of the active user. Second, whether the active user is a frequent traveller, occasional visitor, new to a city or a first time recommender system user. Third, how much is the active user influenced from the activities of other users in the community. The proposed recommender system first learns these factors of an active user and then provides her with a list of personalized recommendations.

The ratings and reviews provided by users at various POIs depict their preferences. Therefore, these ratings and reviews can be utilised to learn the user preferences. The current work addresses two cold start scenarios, 'new user' and 'new city'. The cold start scenarios are presented using Figure 1 and described as below.

I. **The '*new user*' scenario**: A new user does not possess historical data (active user in region-1). Therefore, a recommender system fails to obtain prior knowledge of her preferences or likings. Recommending a set of POIs customized with respect to a 'new' user's preferences is thus a challenge. In this work, we address the problem by exploiting the available ratings of certain trusted users. These trusted users are carefully selected as having the best knowledge of POIs at the concerned region.

II. **The '*new city*' scenario**: A user visits a city for the first time, i.e. the user is not having any historical information at POIs in the city (active user in region-2). It should be noted here that, the new city scenario is different from the user cold start scenario. In a new city scenario, the visiting city is new to the user. However, we have historical data of the user at other regions to identify its interests. Moreover, reviews and ratings from social ties (friends) often influences a user. In this work, preferences of the socially influential users are extensively utilised for POI recommendations.

In general, there exists a set of users who does not fall in the above mentioned categories ('new user' and 'new city'). This set of users consist of both the frequent travellers and the occasional visitors at POIs for a given city. The existing approaches [2, 3, 4, 5] has proposed a single recommendation strategy for both the frequent and occasional travellers. However, a user's preference changes from one region to another based on the contained POIs. Moreover, frequent visits to a region gives a user updated knowledge of the good and the substandard features of the POIs. This knowledge is never gained by an occasional visitor. Therefore, the same recommendation technique aimed at both the type of users is inappropriate. In this work, the frequent travellers and occasional visitors at a region are identified as two different types of users. Separate recommendation schemes are targeted for each of these two type of users.

III. **The '*expert*' user scenario**: The active user at region-3 has 3, 4, 6 and 3 number of visits at various POIs in regions-1, 2, 3 and 4, respectively. It is evident that she is a frequent traveller at region-3, and is believed to have the best knowledge of POIs than the other occasional users at region-3. Such a frequent traveller is termed as an 'expert' of the region. The historical data of the 'expert' is exploited to learn its interests and preferences at various POIs in the region.

IV. **The '*normal*' user scenario**: This represents the general recommendation problem where a user visits a city and requests for a list of customized POIs in and around its current location. The active user at region-4 has 4, 3, 1 and 1 number of visits at various POIs in regions-1, 2, 3 and 4, respectively. She is not a 'new user' as she has historical data from all the four regions. Region-4 is not her 'new city' as she has already visited one POI at region-4. Further, she is an 'expert' of only region-1. Therefore, recommending POIs to such a user is termed as the 'normal' user scenario and the active user as the 'normal' user.

The social connections play a very important role in influencing a users' decision on visiting a POI. Few existing approaches do consider the social influence, but they fail to address the 'new city' problem. Moreover, the existing approaches [2, 3, 4] do not employ distinct recommendation strategies for each of the four category of users, 'new user', 'new city', 'expert' and 'normal' user. The limitations of the existing approaches are mentioned as below.

1. The matrix factorization based approach [11] consider only the user-POI rating data. It fails to utilize the content information and the geographical proximity of the recommended POIs.

2. USG [2] does not consider the reviews provided by users. However, reviews reflect the preferred features of a POI, which in-turn helps in recommendations.

3. USG [2] performs a spatial analysis to identify the distribution of checked-in POIs in a region. Therefore,

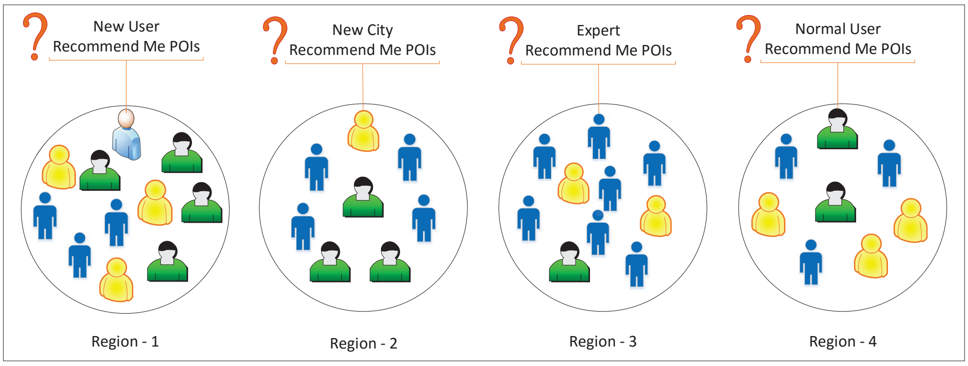


Figure 1: A sample example representing the problems faced by a POI recommender system. The users with a question mark at the top are the active users. Historical or previously recorded data of active users are also shown here. For example, the active user in Region-4 has 4, 3, 1 and 1 historical visit data at Regions-1, 2, 3 and 4, respectively.

the approach solely depends upon the historical data of a user for recommending POIs. The approach fails to address the 'new city' problem, where a user possess no historical data at the 'new city'.

4. CRCF [4] does not consider the proximity of recommended POIs with respect to the active location. However, a recommended POI close to the current location of the querying user generally has a high chance of getting visited.

5. The existing works [2, 3, 4, 11] do not exploit the content based ratings provided by the *experts* of a region. However, it is believed that *experts* can provide unbiased and correct ratings of a POI. Therefore, benchmarking the highly rated features of a POI by these users tend to improve a recommender system.

6. The CRCF approach [4] computes distance between the active location and all POIs. If a POI does not belong to the active region, then its user-POI predicted rating is penalized on basis of the distance factor. The predicted rating of a POI is significantly reduced if it resides far distant from the active location. However, a POI residing within the active region is not penalized with the distance term. A region when considered in terms of a city, consists of a fairly large number of POIs. Hence, the proximity of POIs from the active location also acts as an important factor. Therefore, the geographical influence of POIs within a region should also be considered for recommending the POIs. In this work, all POIs are assigned a score which is weighted by its distance from the active location. This score eventually helps in recommending top-K POIs which are geographically influential.

7. Experiences shared by various users in terms of their activities at a POI, influences an active user to visit the POI. This aspect is not considered in [3, 4, 11] during recommendations. The USG approach [2] performs a friend based collaborative filtering for predicting the rating behavior of the concerned user. However, preferences of the similar users who are not having direct social links, can also help in correlating activities. In this work, we compute social influence of all users on the active user. It is learnt from activities of both friends and similar users.

8. The 'new user' problem is not addressed in the existing works [2, 4, 11]. The proposed work utilises the 'experts' of a region to resolve this problem.

The above mentioned limitations are addressed in the proposed Feature and Region based Recommender System (FRRS). Next, we reproduce certain well-known techniques which are used in FRRS.

4. BACKGROUND OF PROPOSED WORK

A detailed description of certain existing techniques like matrix factorization, friend closeness and pearson correlation co-efficient which are used in our proposed approach is provided in this section.

4.1 Matrix Factorization

Matrix Factorization [10] is a well known model based technique in which first a model is learned and subsequently, the model is used for prediction. The advantage of using MF is that, the rating data is never accessed once the model is build. This helps in working with a large volume of data. Let M be a $U \times F$ user-feature matrix. Each row of M corresponds to active users and corresponding columns depict the features. M_{uf} is the aggregated rating by user u on feature f at various POIs. In the user-feature matrix M, the rating of an unrated feature f for an active user u can be estimated as $\hat{M}_{uf} = y_f^T z_u$, where $y_f \in \mathbb{R}^v$ is the latent v vectors for feature f and $z_u \in \mathbb{R}^v$ is the latent v vectors for user u. First, the two matrices y and z are filled with random values. After each iteration the difference between the predicted rating and the original rating matrix M is collected. The iteration is stopped when a local minimum of difference is obtained. This approach is popularly termed as the gradient descent approach. The squared error between the estimated rating and the original rating can be computed as.

$$E_{uf}^2 = \left(M_{uf} - y_f^T z_u\right)^2 \tag{1}$$

As mentioned in [10], some users tend to rate high for a POI feature and others comparatively rates low. Moreover, there are certain features for which users are more concerned in a POI. For example, the features of a restaurant like food quality and price are of much more importance to customers than the features like aesthetics and branding. Therefore, a rating bias is adjusted with the squared error. Thus the Equation (1) can be rewritten as.

$$E_{uf}^2 = \left(M_{uf} - \mu - b_u - b_f - y_f^T z_u\right)^2 \tag{2}$$

Here μ is the overall average rating on features by all users. b_u and b_f are the average rating by users and on features, respectively. To avoid overfitting, a regularization parameter β is added with the squared error and is represented as follow.

$$\begin{aligned} E_{uf}^2 = &\left(M_{uf} - \mu - b_u - b_f - y_f^T z_u\right)^2 \\ &+ \frac{\beta}{2}\left(\|y_f\|^2 + \|z_u\|^2 + b_u^2 + b_f^2\right) \end{aligned} \tag{3}$$

The goal is to minimize this squared error loss. In this regard, we compute the gradient of the squared error with respect to the two variables, y_f and z_u.

$$\begin{aligned} \frac{\partial E_{uf}^2}{\partial y_f} &= \beta y_f - 2z_u\left(M_{uf} - \mu - b_u - b_f - y_f^T z_u\right) \\ \frac{\partial E_{uf}^2}{\partial z_u} &= \beta z_u - 2y_f\left(M_{uf} - \mu - b_u - b_f - y_f^T z_u\right) \end{aligned} \tag{4}$$

The gradient is next used to estimate the new values for y_f and z_u.

$$\begin{aligned} \hat{y}_f &= y_f + \alpha\left(\beta y_f - 2z_u\left(M_{uf} - \mu - b_u - b_f - y_f^T z_u\right)\right) \\ \hat{z}_u &= z_u + \alpha\left(\beta z_u - 2y_f\left(M_{uf} - \mu - b_u - b_f - y_f^T z_u\right)\right) \end{aligned} \tag{5}$$

This iterative process is continued till the squared error reaches a local minima. The term α is a constant and is used to determine the rate of approaching the minima. Each entry in the user-feature matrix M are found by using the estimated values of y_f and z_u.

The social influence of a friend on an active user can be computed from their 'closeness'. To compute this 'closeness' we utilise the friend closeness approach as depicted in the next section.

4.2 Friend Closeness

The social links or friends tend to share similar activities. This in-turn influences a user to visit a POI. Therefore, computing the 'closeness' of friends on a user is an important parameter for a POI recommender system. The friend closeness mentioned in [16] considers this aspect. It computes the 'closeness' of a social network friend on a user. The 'closeness' between two friends u_1 and u_2 with respect to u_1 can thus be computed as.

$$\gamma_{u_2 u_1} = \delta\frac{|F_1 \cap F_2|}{|F_1 \cup F_2|} + (1-\delta)\frac{|L_1 \cap L_2|}{|L_1 \cup L_2|} \tag{6}$$

where,
F_1, F_2 = friend set of users u_1 and u_2, respectively
L_1, L_2 = checked-in POIs of users u_1 and u_2, respectively
δ = turning parameter which ranges between 0 and 1

Similarity score between the active user and all other users who are not her direct friends can be computed using pearson correlation co-efficient. Details of the similarity score computation technique is mentioned next.

4.3 Pearson Correlation Co-efficient

Pearson Correlation Co-efficient (PCC) [17] is a well known correlation based similarity score computation technique. It is mostly used in finding similarity between a pair of users. PCC between a pair of users is computed based on their ratings on co-rated items. However, in this work, PCC is used to find similarity between two users by exploiting the ratings given by the concerned users on the co-rated features of POIs. Thus the feature based similarity score between a pair of users u_1 and u_2 can thus be computed as.

$$\lambda_{u_2 u_1} = \frac{\sum_{i \in M_f}(r_{u_1 i} - \bar{r}_{u_1})(r_{u_2 i} - \bar{r}_{u_2})}{\sqrt{\sum_{i \in M_f}(r_{u_1 i} - \bar{r}_{u_1})^2}\sqrt{\sum_{i \in M_f}(r_{u_2 i} - \bar{r}_{u_2})^2}} \tag{7}$$

where,
M_f = Common features of u_1 and u_2

Utilizing the above three techniques we devise the Feature and Region based Recommender System. The following section describes the working of each module of the proposed FRRS framework.

5. PROPOSED FRAMEWORK

In this section, we describe the detailed methodology of the proposed Feature and Region based Recommender System (FRRS). The framework of the proposed approach is depicted in Figure 2. Detailed description of each module is presented next.

5.1 Modelling User Preferences

In general, people always prefer significant places. The significance of a POI to a user is mostly driven by its liking for features or aspects of a POI. Therefore, identifying a user's interest or liking on certain features of POIs is the primary task for any recommender system. This section explores the review and rating data to identify the inherent interests or affections of each user. People often rate and post a review on POIs, either through website or mobile applications provided by Facebook, Twitter, Foursquare, Instagram, Yelp, etc. Ratings given by users always reflect their interest on a POI. The reviews on the other hand depicts the good and bad features of a POI. Incidentally, it has also been noticed that users generally post reviews on certain features of POIs which are more significant to them. In this work, we exploit the reviews on POIs by a user to learn his/her interests. We apply vector space model to extract features from review data, as described in [4, 18]. The rating given along with the review, portrays the dimension of liking or disliking a feature. If a feature is rated within a scale of 1 to 5, then a feature dimension say 5 represents the most favoured feature, whereas, a value of 1 represents the disliked feature. There can be two major ways in which the user can provide a rating at a POI. First, a user can explicitly select a feature while rating a POI. He may select

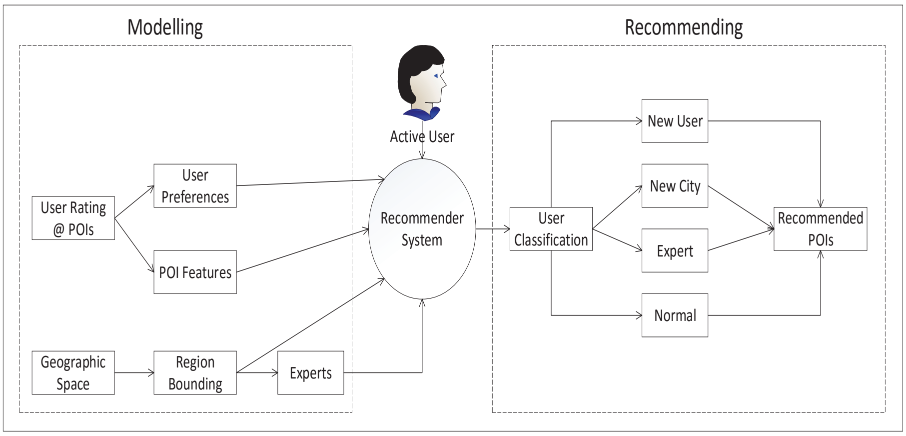


Figure 2: The proposed framework of FRRS for recommending top-K POIs to a user.

cleanliness as a feature of a hotel in New York and provide a good rating of 5. Similarly, he may also select room service as a feature of the same hotel and rank it with a poor rating of 1. Second, the users can rate a POI without explicitly mentioning a feature. This type of rating is known as implicit rating. For an implicit rating, we consider the same rating to all the features of the POI. For example, if a user implicitly rates a coffee shop at 4, then we consider the rating at the associated features like quality, service, price, etc. also as 4.

We generate a user-feature matrix where each feature denotes the subject of interest and its dimension of liking for the users. The final rating by u_1 on a feature f_1 is the average rating given by u_1 on f_1 of all the POIs in which it has rated. The user-feature rating matrix consists of ratings either given explicitly or implicitly by the user. From this matrix, we identify the set of top-W features (having the highest aggregated rating) as the preferred features of a user. However, there exist many missing ratings for certain features. To address this problem, we employ the basic matrix factorization (MF) technique [10]. Details of this technique is elaborated in Section 4.1. Each row-wise entry in the user-feature matrix M gives a clear understanding of the preferred features of a user.

5.2 Modelling Features of POIs

The reviews provided by users on a POI mostly focus on certain specific features. The complete set of features consist of both the most favoured features and also the much disliked ones. Similar to the approach for modelling user preferences, a feature-POI matrix is generated. It can be noted here that, features of a POI vary on basis of the category of POI. Like, a stadium can have a feature like outfield, whereas a museum can never have such a feature. Similarly, demonstration can be a feature at a museum but can never exist for a stadium. However, certain features like car parking, ticket rate, etc. falls for both the POIs. Therefore, the features which are extracted from the set of reviews at a particular POI are only considered as the bag of features representing the POI. Similar to the previous case, we utilize the aggregated ratings (both explicit and implicit) given by users for a feature of a POI. The feature value can be zero if the feature is absent in a POI. To select a set of marque features of a POI, the top-v features are chosen having high ratings. The ratings are normalized within a scale of 1 to 5, where 5 represents a good rating and 1 means a substandard feature. A feature-POI matrix N is generated with each row representing a feature and the columns representing POIs. An entry N_{fp} depicts the average rating provided by users for the feature f at POI p. Therefore, each column vector of N depicts the popular features of a POI. The obtained user and POI features from M and N, respectively are meticulously used in the recommendation part of FRRS.

After the user preferences and features of POIs are learned, the geographic space needs to be clustered into regions. This is required to identify the regions within which the POIs exists. The next section describes the technique employed for region bounding.

5.3 Region Bounding

The geographical area over which we apply the proposed POI recommender system needs to be divided into regions to understand the locality of a POI. Identifying the region or locality of a POI is very important in context of an efficient recommender. This is due to the fact that, people requesting for a recommended list generally likes the results to be within a feasible distance so that they can be visited. If a user requests for a museum from New Delhi, he will always desire to get a list of museums in around the current location. Even though the British Museum at London may arguably be of high rating than a local museum at New Delhi. Hence, identifying the locality of a POI is a necessity to provide efficient recommendations. The k-means clustering algorithm is used to divide the geographic space into regions and thus locate each available POIs into a specific region.

Subsequently, for each region we classify the frequent users and the occasional visitors. It is expected that the frequent visitors will have better knowledge of the good and also the substandard features of the POIs in a region. In this work, the user who is having more number of rating and review activities at a region (say, R) is classified as the frequent traveller at R. All other visitors having less number of review-rating activities at R are thus marked as the occasional visitors. As mentioned in Section 3, the frequent and occasional travellers are termed as 'expert' and 'normal' users, respectively. Ratings provided by the *experts* of a region is given a higher preference than other users while recommending POIs to a 'new user'. The next section depicts how the bounded regions and user features are utilised to predict rating of a POI.

5.4 Predicting Rating of a POI by a User

To recommend a set of POIs to a user, the recommender must learn how likely the user will visit the POIs. The different variations of the collaborative filtering like the item based, user based and friend based approaches have been employed in this work. Let U be the set of all user $\{u_1, u_2, ..., u_m\}$, G be the set of all regions $\{g_1, g_2, ..., g_n\}$ and P be the set of all POIs $\{p_1, p_2, ..., p_o\}$. The task of estimating rating ER for a user u_1 at a POI p_1 located in a region g_1 faces four different scenarios. In the rest of this section, we broadly explain each of these scenarios and also state the technique employed by our approach to estimate ER for u_1 at p_1.

5.4.1 u_1 is a 'new user'.

As mentioned earlier in Section 3, a 'new user' has no historical data. Therefore, a recommender system fails to learn the new user's preferences. To address this issue, our proposed recommender utilises the ratings given by the trusted *experts* of the regions. The aggregated rating provided by the *experts* E of region g_1 is used as the estimated rating ER on p_1 for the 'new user' u_1.

$$ER_{p_1}^{u_1} = \frac{1}{|E|} \sum_{e \in E} r_{ep_1} \tag{8}$$

5.4.2 g_1 is a 'new city' to u_1.

According to the 'new city' problem stated in Section 3, u_1 has not rated any POI within region g_1. This problem can be solved by considering the collaborative rating given by the most influential users of u_1. In this regard, we first consider the friends as they often share similar interests [19]. This aspect can be exploited in the current work. For the connected social links or friends we find similarity by utilising the friend closeness. The social influence of a friend u_2 on the active user u_1 is determined from the set of mutual friends and the common POIs visited by them. Details of the computation technique for 'closeness' between friends is mentioned in Section 4.2. The friend closeness scores γ between the active user u_1 and all its friends are considered as their social influence on u_1.

Incidentally, this approach can lead to a limited coverage problem. This is due to the fact that, for a particular POI p_1 we may get zero or very less number of friends of u_1 who have actually visited it. For example, a new or uncommon POI may not attract a large number of users and thus the set of friends who have visited such a POI will be less. However, there is always a chance of having users with similar taste as that of u_1. The social network users having similar taste or feature set as that of the active user, without any direct social links between them is termed as similar users. The similar users are often found to influence the activities of an active user [20, 21]. The similarity between two users (not friends) can be found by computing a correlation score between their preferred features as depicted in their respective feature matrix M. To compute user similarity score, the popular Pearson Correlation Co-efficient (PCC) measure is used. The detailed description of the feature based correlation score between a pair of users using PCC is depicted in Section 4.3. The positive correlation scores λ between the active user u_1 and all other users are considered as their social influence on u_1. This approach helps us to identify the socially influential users which consists of both the friends and the similar users. Subsequently, the top-I users who visited p_1 having high social influence on the active user u_1 are selected. Finally, the estimated rating ER at p_1 for u_1 with $I = \{i_1, i_2, ..., i_d\}$ influential users is computed as.

$$ER_{p_1}^{u_1} = \frac{1}{|I|} \left[r_{i_1 p_1} + \sum_{j=2}^{d} \frac{r_{i_j p_1}}{\log_2(j)} \right] \tag{9}$$

where,
$r_{i_1 p_1}$ = rating given to p_1 by user i_1 at position 1 of the ranked list I
d = number of users in I

5.4.3 u_1 is an 'expert' at g_1.

As already mentioned in Section 5.3, a user is considered as an *expert* of a region if she is having maximum number of ratings at the region. Moreover, it is also expected that the *experts* has sufficient knowledge for distinguishing between the good and the substandard features of a POI. Here we apply the standard Item based Collaborative Filtering (IbCF) approach. In the present context, the items refer to POIs, and hence we term this technique as the POI-based Collaborative Filtering (PbCF). The PbCF is a neighborhood based approach for predicting the rating given by a user to a target POI. In the context of this work, the PbCF approach first finds the similarity score between p_1 and all other POIs already rated by u_1. The Pearson Correlation Co-efficient (PCC) similarity measure is used to compute the similarity score between p_1 and all POIs located in region g_1. Here the precomputed feature-POI rating matrix N is used to find the average rating on a feature of a POI. Subsequently, the estimated rating ER on p_1 by u_1 is computed as follow.

$$ER_{p_1}^{u_1} = \bar{r}_{p_1} + \frac{\sum_{x \in X} sim(p_1, p_x) * (r_{u_1 x} - \bar{r}_x)}{\sum_{x \in X} |sim(p_1, p_x)|} \tag{10}$$

where,
$ER_{p_1}^{u_1}$ = estimated rating at p_1 for u_1
$\bar{r}_{p_1}$ = average rating at p_1
X = set of all POIs rated by u_1
$sim(p_1, p_x)$ = similarity score between POIs p_1 and p_x
$r_{u_1 x}$ = rating at POI p_x by u_1

The PbCF approach is vulnerable for the new user cold start problem. This is due to the fact that, a new user has no POI ratings and thus it is very difficult to identify the set of POIs with similar taste. Incidentally, in this scenario the active user u_1 is an *expert* of region g_1. Therefore, u_1 will always have sufficient ratings to perform PbCF.

5.4.4 u_1 is a 'normal' user at g_1.

This particular scenario deals with a situation when u_1 has certain number of historical details at g_1, although not enough to consider him as an *expert* of g_1. The pattern of rating POIs by u_1 can be obtained from its published reviews and ratings at various POIs located within g_1. Moreover, the friends and similar users can influence u_1 to visit POI p_1. Therefore, we approach this scenario by proposing a hybrid technique which combines both the POI-based collaborative filtering and the impact of direct and indirect social links. The estimated rating ER on p_1 for u_1 in this scenario is computed as follow.

$$ER_{p_1}^{u_1} = \tfrac{1}{2}\left[ER1_{p_1}^{u_1} + ER2_{p_1}^{u_1}\right] \tag{11}$$

where,
$ER1_{p_1}^{u_1}$ = Estimated rating on p_1 for u_1 from Equation (10)

$ER2_{p_1}^{u_1}$ = Estimated rating on p_1 for u_1 from Equation (9)

It can be noted here that, the major limitation of collaborative filtering is the limited knowledge of review and rating data. Although u_1 is not an *expert* of g_1, still it has enough review and rating data to perform a collaborative filtering technique. Moreover, combining the impact of friends and similar users enhance the chance of recommending personalized POIs.

5.5 Recommending top-K POIs

The procedure of recommending the final list of top-K POIs is described in this section. A POI recommender is widely accepted if it provides POIs nearby to the active location. Moreover, the geographical influence of proximal POIs has more impact on recommendations than the social influences [2]. A simple approach is often preferred, which computes the distance of POIs from the active user's location and recommend the top-K POIs closest to it. However, this approach does not agree with the feasibility of liking the recommendations, as this can select a POI with negative rating in the recommended list. Therefore, it is necessary to combine both the rating of a POI and it's distance from the active user to finalize the recommendations. In this regard, we introduce a term Score for each POI with respect to a user. Score of a POI p_1 with respect to a user u_1 is computed as.

$$Score_{p_1}^{u_1} = \frac{ER_{p_1}^{u_1}}{dist(u_1, p_1)} \tag{12}$$

Here, $dist(u_1, p_1)$ is the geographical distance between the current location of u_1 and the POI p_1. This Score of each POI is thus used to perform a personalized ranking of POIs with respect to the active user. The top-K POIs having high Scores are considered for the final recommendation. Details of our proposed approach is depicted in Algorithm 1.

6. EXPERIMENTS

We performed experiments on the real-world Yelp dataset [1]. All experiments are implemented using Matlab. The proposed POI recommendation approach is compared with a baseline approach [5] and three existing approaches [2, 3, 4]. The Probabilistic Matrix Factorization (PMF) is considered as the baseline approach in this work.

[1] *https : //www.yelp.com/dataset_challenge*

Algorithm 1: Feature-Region based Recommender System (FRRS)

Input: D: Yelp Dataset
Result: RL: Recommended list of top-K POIs
Data:
ER = Estimated rating of a user u_1 at a POI p_1 located in a region g_1
SI = Social influence

Generate user-POI rating matrix S
Extract features F from user reviews
Generate user-feature rating matrix M from S
Generate feature-POI rating matrix N from S
Perform matrix factorization on M
Cluster the geographical space into small regions by k-means clustering
A=extract review and rating data of u_1 from D
B=extract review and rating data of u_1 at g_1 from A
C=region for which u_1 is an 'expert'
if *A has no data* **then**
    u_1 is a 'new user'
    ER = average ratings at POIs by the 'experts' of g_1
else
    if *B has no data* **then**
        u_1 is in a 'new city'
        Compute SI of each friend on u_1 using friend closeness
        Compute SI of other users (not friends) on u_1 using PCC
        I = all users arranged on basis of SI on u_1 from high to low
        Select top-I most influential users for u_1 who have visited p_1
        ER = compute from ratings of the selected I users
    else
        if *C and g_1 are same region* **then**
            u_1 is an 'expert'
            ER = POI based collaborative filtering on reviews of u_1 at g_1
        else
            u_1 is a 'normal' user
            ER = combine ratings by 'experts', friends and similar users of u_1
        end
    end
end
Compute *Score* for each POI with respect to u_1
RL = Recommend top-K POIs with high *Score*
return RL

6.1 Dataset Description

Yelp provides crowd-sourced reviews on local business. The publicly available Yelp dataset has been used in this work to report the experiments. It consists of 45,981 users, 2,29,906 ratings and 11,537 POIs. The reviews are first preprocessed to remove the stop words. Subsequently, the infrequent words (words occurring in less than 100 reviews) are removed. The remaining set of words from the reviews (8519 words) are considered as the feature set.

6.2 Evaluation Technique

In this work, four different scenarios in POI recommendation has been considered (Section 3). We devise a strategy which can evaluate the baseline, existing and our approach for the two cold start scenarios and the 'expert', 'normal' user recommendation scenarios. Further, for each of these scenarios we test the error in rating prediction and accuracy of the recommendations. Below we mention the strategy used to generate the test datasets for evaluating the existing and proposed recommendation approaches.

I. **$u1$ is a '*new user*'.**
Users having less than 20 ratings are considered as 'new user' and their historical data is used for testing the 'new user' cold start scenario.

II. **g_1 is a '*new city*' to u_1.**
For each user, we first identify the region with least number of reviews. The historical data of each user at these regions are selected for testing the 'new city' cold start scenario.

III. **u_1 is an *expert* at g_1.**
We randomly select 10% ratings of every user from the region in which they are 'experts', and this dataset is used for testing the 'expert' user scenario.

IV. **u_1 is a '*normal*' user at g_1.**
We randomly select 10% ratings performed by every user at regions in which they are not 'experts', and this dataset is used for testing the 'normal' user scenario.

The actual rating provided by the users to the POIs in the test dataset acts as the ground truth for all the above cases. It can also be noted here that, while computing *Score* (Equation (12)) the current location of the active user is selected as the region centre in which the POI belongs. The error in rating prediction is estimated using the Mean Absolute Error metric. To evaluate the accuracy of the recommended list of POIs, we adopt a similar technique as applied in [4].

In our experimentation, the dimension of user factor model and feature factor model is selected at 10. The parameters α and β used during matrix factorization is fixed at 0.0001 and 0.01, respectively (Section 4.1). While computing friend closeness, the turning parameter δ was set at 0.3 with a high weightage to the common POIs visited than the number of common friends (Section 4.2). The geographical space is clustered into 10 regions i.e. k=10 (Section 5.3).

6.3 Results and Analysis

In this section, we first report the results and then provide a detailed analysis on them.

6.3.1 Error in Rating Prediction

The Mean Absolute Error (MAE) has been widely used as a statistical accuracy metric for measuring the error in rating prediction for various recommender systems [7]. It estimates the average deviation of the predicted rating from the actual rating of a POI by a user. The Mean Absolute Error (MAE) in predicting ratings for all the active users U can be represented as.

$$MAE = \frac{1}{|U|} \sum\nolimits_{i \in U} \frac{1}{|G_i|} \sum\nolimits_{j \in G_i} \left| ER_j^i - r_{ij} \right| \quad (13)$$

Table 1: MAE of various approaches in dealing with the four different scenarios mentioned in Section 3.

	New User	New City	Expert	Normal
PMF	1.415	0.886	0.576	0.769
USG	1.328	0.882	0.574	0.767
LCARS	1.326	0.879	0.574	0.763
CRCF	1.211	0.875	0.571	0.762
FRRS	**0.954**	**0.851**	**0.566**	**0.757**

where, G_i is the set of all POIs in the test dataset for user i. ER_j^i is the estimated rating on POI j for user i using various approaches, and r_{ij} is the actual rating by user i on POI j.

Table 1 shows the MAE in predicting user ratings to POIs using various approaches. The results categorically depict their performances in dealing with the four scenarios already mentioned in Section 3. It can be observed from the results that the compared approaches produce nearly similar results for the 'expert' and 'normal' recommendation scenarios. However, for cold start scenarios like 'new city' and 'new user' the proposed FRRS produces significantly less error rate than the others.

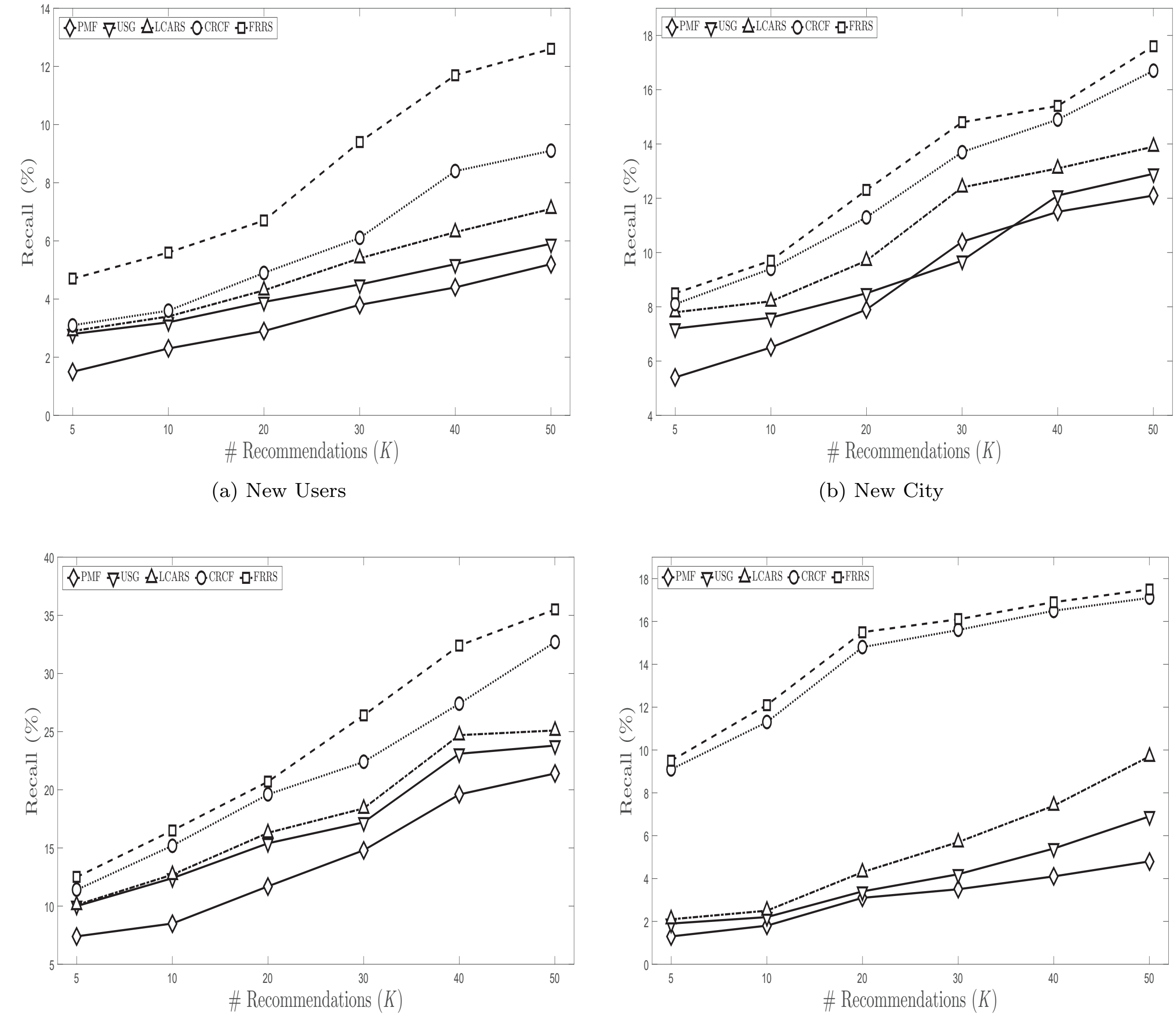
6.3.2 Accuracy of Recommendation

During experimentation, we recommend relevant POIs for each user. The accuracy of recommendation is estimated on basis of whether the POIs in the test dataset occurred in the recommended list. In this regard, we employ the *Recall* metric [4]. It can be represented as,

$$Recall = \frac{\#Hit}{\#Ground\,truths} \quad (14)$$

Hit is the number of POIs in the test dataset which are also observed in the recommended lists. $Ground\,truths$ is the number of POIs in the test dataset.

Figure 3 (a-d) shows the performance comparison of different recommendation approaches for specific scenarios. The 'new user' scenario is not specifically considered in CRCF, LCARS and USG, whereas our approach utilises the ratings given by *experts* to address this issue. A notable improvement in recall using our approach is visible from the obtained results. In a 'new city' scenario, our approach considers the ratings of influential users. This improves performance of our approach than other existing approaches which fails to consider both the friends and similar users. All the approaches are observed to have comparatively high recall for the third scenario, where an *expert* of a region is provided with the recommended list. This is mostly due to the fact that an *expert* has a dense number of historical data at their respective home regions. This reduces the possibility of a cold start scenario and similarly increases knowledge of a user's likings. In 'normal' scenario, our proposed approach produces better results than CRCF, LCARS, USG, and significantly higher than the baseline PMF. The obtained results show that our approach clearly outperforms others. With increase in the number of recommendations, the possibility of a Hit increases and this is clearly reflected from the obtained results for all the approaches.



(a) New Users (b) New City

(c) Expert Users (d) Normal Users

Figure 3: Recall of the compared approaches in dealing with the four different scenarios mentioned in Section 3 is presented here. The number of recommendations K in horizontal axis is varied from 5 to 50.

7. CONCLUSION

This work studies the preference of users and their geographical locality to recommend POIs. First, a matrix factorization based technique is employed to learn the preferences of users from their reviews at various POIs. Social influence on a user and the active location are believed to be the two most important factors for an effective recommendation. The proposed work computes the impact on a user from its connected social links or friends, and also from the users following similar activities. Moreover, a score is assigned to every POI by adjusting the estimated rating at it using a distance parameter. The top-K POIs with high score are recommended by our approach. Moreover, the proposed recommender system addresses two cold start scenarios, 'new user' and 'new city'. General users are also classified into frequent and occasional travellers. Distinct recommendation strategies are also proposed for each of the two types of travellers. Along with this, there still exists a scenario where a POI 'new' to the recommender system never gets recommended. In this scenario, the recommender system fails to learn the POI features. Therefore, even if the 'new' POI is relevant to a user, still it never gets recommended. This problem can be termed as the 'new POI' problem for a recommender system. In future, we plan to extend this work to solve the 'new POI' problem.

8. REFERENCES

[1] Gediminas Adomavicius and Alexander 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*, 17(6):734–749, 2005.

[2] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun 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*, pages 325–334. ACM, 2011.

[3] Hongzhi Yin, Yizhou Sun, Bin Cui, Zhiting Hu, and Ling Chen. Lcars: a location-content-aware recommender system. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge discovery and data mining*, pages 221–229. ACM, 2013.

[4] Chenyi Zhang and Ke Wang. Poi recommendation through cross-region collaborative filtering. *Knowledge and Information Systems*, 46(2):369–387, 2016.

[5] Ruslan Salakhutdinov and Andriy Mnih. Probabilistic matrix factorization. In *Proceedings of the 21st Conference on Neural Information Processing Systems*, pages 1–8, 2011.

[6] Wen-Yen Chen, Jon-Chyuan Chu, Junyi Luan, Hongjie Bai, Yi Wang, and Edward Y Chang. Collaborative filtering for orkut communities: discovery of user latent behavior. In *Proceedings of the 18th International Conference on World wide web*, pages 681–690. ACM, 2009.

[7] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, pages 285–295. ACM, 2001.

[8] Mohsen Jamali and Martin Ester. Trustwalker: a random walk model for combining trust-based and item-based recommendation. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge discovery and data mining*, pages 397–406. ACM, 2009.

[9] Xin Jin, Yanzan Zhou, and Bamshad Mobasher. A maximum entropy web recommendation system: combining collaborative and content features. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge discovery in data mining*, pages 612–617. ACM, 2005.

[10] Yehuda Koren, Robert Bell, Chris Volinsky, et al. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[11] Nathan N Liu, Xiangrui Meng, Chao Liu, and Qiang Yang. Wisdom of the better few: cold start recommendation via representative based rating elicitation. In *Proceedings of the 5th ACM Conference on Recommender systems*, pages 37–44. ACM, 2011.

[12] Badrul M Sarwar, George Karypis, Joseph A Konstan, and John T Riedl. Application of dimensionality reduction in recommender system–a case study. In *ACM WEBKDD WORKSHOP*, 2000.

[13] Julian McAuley and Jure Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM Conference on Recommender systems*, pages 165–172. ACM, 2013.

[14] Bracha Shapira, Francesco Ricci, Paul B Kantor, and Lior Rokach. *Recommender Systems Handbook.* Springer, 2011.

[15] Vincent W Zheng, Yu Zheng, Xing Xie, and Qiang Yang. Collaborative location and activity recommendations with gps history data. In *Proceedings of the 19th International Conference on World wide web*, pages 1029–1038. ACM, 2010.

[16] Zheng Huo, Xiaofeng Meng, and Rui Zhang. Feel free to check-in: Privacy alert against hidden location inference attacks in geosns. In *International Conference on Database Systems for Advanced Applications*, pages 377–391. Springer, 2013.

[17] Christian Desrosiers and George Karypis. A comprehensive survey of neighborhood-based recommendation methods. In *Recommender systems handbook*, pages 107–144. Springer, 2011.

[18] Chenyi Zhang, Ke Wang, Ee-peng Lim, Qinneng Xu, Jianling Sun, and Hongkun Yu. Are features equally representative? a feature-centric recommendation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pages 389–395. AAAI Press, 2015.

[19] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.

[20] Philip Bonhard, Clare Harries, John McCarthy, and M Angela Sasse. Accounting for taste: using profile similarity to improve recommender systems. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1057–1066. ACM, 2006.

[21] Philip Bonhard, M Angela Sasse, and Clare Harries. The devil you know knows best: how online recommendations can benefit from social networking. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1*, pages 77–86. British Computer Society, 2007.