New Location Recommendation Technique on Social Network

Sutarat Choenaksorn
Department of Mathematics and Computer Science,
Faculty of Science, Chulalongkorn University
Bangkok, Thailand
schoenaksorn@gmail.com

Saranya Maneeroj
Department of Mathematics and Computer Science,
Faculty of Science, Chulalongkorn University
Bangkok, Thailand
saranya.m@chula.ac.th

ABSTRACT
With the availability of current modern technologies, decisions making in an everyday life can be assist in many different ways. Many researches in the past decade has studied about recommendation systems. Recommendation systems can base on different variables with location-based services is one of the more interesting factor to a recommendation system. Recommendations on Location based Network is a service for assisting people to locate locations of their interests. A large number of recorded checked-in histories was gathered to make the prediction according to the desired preferences of each user. Furthermore, determinations have shown a social relationship leading to availability of information will assist in making better recommendations based on the locations. Recently, the recommendation system on location-based domain usually combines either content-based technique and collaborative technique, or collaborative technique and social-based techniques. It is difficult to find the way to combine those three techniques. So there is no research that combine those techniques on location-based recommendation system. This study proposes a new method that combines content-based technique, collaborative technique, and social-based techniques; to produce more efficient result results than location-based RS methods. The evaluation results show that the proposed method provide higher accuracy and coverage than two current location methods by measuring with the Normalized Discounted Cumulative Gain (NDCG) and coverage matrix.

CCS Concepts
• **Information systems → Recommender systems**

Keywords
recommender system (RS); collaborative filtering (CF); content-based filtering (CBF); social filtering

1. INTRODUCTION
The Recommender System (RS) is a system, which uses to recommend the interesting items to an active user. The recommendation has several techniques for seeking the items, e.g. content-based filtering (CBF), collaborative filtering (CF), hybrid recommender system (Hybrid System) [1,2]. However, each of the techniques has the difference method for the recommendation. The CBF recommends the new items that user might prefer by learning from the user's behavior in the past. The CF [3] relies on the neighbor's opinion to recommend the new interesting items. The neighbors are the set of users who have the most similar taste with the active user. The Hybrid System is a technique that combines more than one techniques by fixing the drawback of each technique to create a better performance.

With the popularity of social network, location data are extensively used such as sharing the location, checking in the location. Location data is a raw data that relate to the longitude and latitude referring to locations. Moreover, there is also relationship on social networks that are used widely. The Relationship between a pair of user in Social network can be called in many names for example in Facebook, we call friends or in Twitter, we call followers. In the relationship it can be shown that some group of people affect the another group who have similar life styles. The person who can create such impact can be celebrities or persons who well-known by numerous of followers. Therefore, this impact level should be used as another factor in prediction process. Many researchers have proposed social filtering technique in order to search these impacts.

In currently, there are several the RS, that concentrate upon the location domain. For instance, in work of [4], they use the CF technique to seek the relationship of location popularity, user similarity and location similarity from past check-in information of user, and show three relationships into three sub-matrixes. After that, they are combined into a matrix and used to analyse for generating popularity value to show how much is user will visit each location**.** In work of [5], they use the CF technique to find the travel intent of each review by average the intention value then plot the graph that contain node, link and edge. Each place represents by node and intention value contain in node. Link represent a transportation type between node (place) such as road, public transportation. Weight represent a distance between node. The graph act as the guideline for going to the place along with visiting objective and how to get there. There are also many researchers who are interested in using a hybrid system in a domain location for example, in work of [6] and [7], the interesting locations are recommended by combined with the CBF technique and the CF technique. The first step, the user preference is sought by the CBF technique. Subsequently, the user preference of the users is used to calculate the similarity between them by using the CF technique and ranks the interesting location in the last step. In work of [8], the recommended locations are created by the combination of social influence and geographical influence. The first step uses CF technique and social filtering to calculate for the user similarity and social similarity, respectively, after that multiplying the both results to find combinative similarity value (A), then find social influence value by weighted

ICISS '18, April 27–29, 2018, Jeju, Republic of Korea

ACM ISBN 978-1-4503-6421-8/18/04$15.00
https://doi.org/10.1145/3209914.3209933

(A) with place's indicator of each user. Subsequently step, the users' physical check-in behaviour is used to create the geographical influence. The last step, the value of geographical influence and the social influence will be combined, and the location recommendation for user will be provided.

Previous research has combined the CF technique and the CBF technique together, but there is no research that combines three techniques for recommending location. In order to create the better recommendation performance, this paper proposes a new method that combines the CBF, CF, and also the social filtering technique. The proposed method demonstrates the finer accuracy outcome while compares with two methods, i.e. the first method combines the CF and the CBF technique the second method combines the CF and the social filtering technique by using the Normalized Discounted Cumulative Gain (NDCG) approach to evaluate the recommendation accuracy.

The paper presents the related works in section II, section III presents a new method for recommendation, section IV presents an experiments, and the section V which concludes this paper.

2. RELATED WORK

2.1 CF on Location Recommend

Guo *et al.* [4] propose a method for introducing attractions for individual travelers. The first step creates a bi-part graph to show the link between the two nodes: the user and the location. The weight of the link is based on the frequency information that the user checks in each place. Second step uses the CF technique to calculate three types of relationships. These three relationships are recorded in form of three sub-matrixes which are the matrix that shows the relationship between users and locations (location popularity), the matrix that shows the relationship between a pair of users (user similarity), and the matrix that shows the relationship between a pair of locations (location similarity). These three relationships use the data from the bi-part graph is then compute to seek these the relationship into the from of three sub-matrixes are combined into one big matrix. The third step, that matrix is then that calculated and analysed to get the popularity value which show how much is user will visit each location. When the analysis completed, we get the user probability to visit each location. The last step, that popularity value of locations are ranked for users' recommendation.

Kim et al. [5] propose a method for introducing attractions for individuals by analyzing the scope of intent. It is divided into 8 groups: Business and professional, Eating out, Education and training, Health and medical care, Holiday, leisure, and recreation, Religion and pilgrimages, Shopping, and Socializing (friends and family). For the first step uses the CF technique to seek the intention's similarity between each review and intentional group by using the comparison of word vector between review and intention group. The second step find the visiting target score of each place by using the similarity between review and intention group of all review of each location to calculate for the average and normalize in between "0-1**.** In the last step uses the obtained data to present in form of weighted graph which contain nodes, links and weight. The node refers to the location. In addition, each node also shows the visiting target score of each place from the second step, the line refers to type of transportation between nodes (locations) e.g. roads, public transportation and etc., and the weight refers to the distance between nodes. The weight graph is used as the decision making information for users to decide that each location is suit to visit in which way.

2.2 CBF and CF on Location Recommend

Rengith *et al.* [6], they present a personalized travel recommender model. This model has two processes. Firstly, they use CBF technique to create the user preference. Then they use the CF technique to calculate the potential target location based on their preference. Finally, they map the target location and the user physical check-in to find the location that is nearest the user.

Yu *et al.* [7], they propose a method for introducing attractions for individual by introduce recommended sightseeing package. It can divide to 6 period and each period can find recommended place by following steps: First step uses the CBF technique to calculate the user preference of each location that represented as a user profile. Second step calculates the average popularity of each location in each month. Third step uses the CF technique to calculate the relationship between the users by comparing the user profile derived from the first step. Fourth step calculates the predicted rating for each user on each location by using the weighted average on user similarity and location popularity**.** The last step, for sightseeing introduction by mapping the high predicted rating with user's location then choose the nearest location and add to the package.

2.3 CF and Social Filtering on Location Recommend

Zhang *et al.* [8] propose a method for introducing individual attractions by using Social and Geographical Fusing Model (SGFM), which is created by social influence and geographical influence value. For the creation of the social influence, it can be divided into these following steps: the first step is searching user similarity by using CF technique to seek the relationship between user and the other users. The second step is searching similarity among users using social filtering technique from their mutual friends. More mutual friends mean they are very similar. In the other hand, less or zero mutual friends mean less or no similarity on those users. The third step is calculating combinative similarity value (A) by multiplying user similarity value with social similarity value. The fourth step is calculating social influence value of each user by weighted (A) with place's indicator of each user. The indicator means that the user who used to go to that place is 1. On the other hand, 0 is the user who never go. Next step for geographical influence value searching, it can be calculated from the user's check-in behaviour. The next step is the obtain-from-calculation data analysis step, which is the multiplying between the social influence and geographical influence to seek the recommended location for the users. Proposed Methods.

Previous researchers have combined either the CF technique and the CBF technique or the CF technique and social filtering together, but there is no research that combines those three techniques for recommending location that assume to provide more accurate prediction result.

3. PROPOSED METHOD

This paper, in order to create the better recommendation performance, this paper proposes a new method that combines the CBF, CF, and also the social filtering technique following instruction. It consists of four steps.

3.1 Create User Preference

The User Preference is the value that shows the relationship between an individual user and a location. This relationship is derived from the user's visit to that location and check in at the place. A large number of recorded history records are used to seek the preference of each user. For calculation of the preference of each user, the method bring the number of times that user u checks in to each location, and normalize with the total number of user u that checks in, which makes it possible to calculate the user preference of each location. Near-preference 1indicates the user's preference with the location. On the other hand, close proximity value 0indicates the user's preference with the location less. We can calculate the user preference by the equation below.

$$F_o^u = \frac{VC_{(u,o)}}{VC_{(u)}} \quad (1)$$

Where F_o^u means the preference of the user *u* toward the location *o*, $VC_{(u,o)}$ means number of user *u* checking time to the location *o*, $VC_{(u)}$ means the total number of checking in time that user *u* checked in.

After calculation of user preference u at all the locations, the vector can be used to represent user preference profile as equation 2.

$$F^u = [F_{o1}^u, F_{o2}^u, \dots, F_{on}^u] \quad (2)$$

WhereF^u represents the vector of user preference u for all locations, F_o^u is the value of user preference profile with place *o*, which is calculated by the equation (1)1 ,2 , ..., *n* denote the number of locations from order 1 to order *n*.

3.2 Find User Similarity

After ability to calculate the satisfaction of users in every place, it can precede the process of calculating the similarity between the users. The value represents the relationship between two users in a manner that represents a similar personal preference. For calculation of the similarity between the users, this can be done by using cosine similarity on the vector of user preference profile of user *u* and *v* that is derived from equation (2). The similarity between users that are close to 1 indicates that two users have a preference for similar places. On the other hand, the user proximity value 0 indicates that two users have a preference for a different location. We can calculate the similarity between the users by the equation below.

$$Sim_{(u,v)} = \frac{F^u * F^v}{\sqrt{F^{u^2}} * \sqrt{F^{v^2}}} \quad (3)$$

Where $Sim_{(u,v)}$ means the similarity between the user *u* and the user *v*, F^u and F^v vector of user preference profile for all locations of user *u* and *v*. At this step, a similarity of user and target user can find by using CF technique.

3.3 Find Social impact of Each user

The social impact is the value that represents the influence of user *u* that affects to other people in the online society. In this paper for the social impact of each person, we refer to the fame. It can see from the number of the user's followers. Therefore, it means that a person who has fame or reliability, they must have lots of followers. Thus, calculating is calculated by the number of followers of user *u* normalizes with the highest number of followers of all users in accordance with equation (4). The social impact that closes to 1 indicates user *u* has a huge impact to people on social networks. On the other hand, the value of social impact that closes to 0, user *u* has less impact to people on the social network.

$$IP_u = \frac{FR_u}{Max(FR)} \quad (4)$$

Where IP_u means the value of social impact on users u on the social online, FR_u the follower number of user u, $Max(FR)$ the maximum number of followers of all users. At this step, the impact of users can find by using social filtering technique.

3.4 Calculating Predicted Rating of Locations

The probability of *u* will visit each place *o* can be calculated by multiplying the result of step 2 with result of step 3 and weighted by the number of times that other user *v* have checked in the place *o*. This makes the importance of the location of each user has a different away. The ratings are between 0-1. The calculated ratings will be ranked in descending order. The rating of the nearest 1 will be ranked at first to use to guide users. So we can calculate the rating by using the equation as follows:

$$R_{(u,o)} = \frac{\sum_{v \in U} Sim_{(u,v)} IP_v Freq_{(v,o)}}{\sum_{v \in U} Sim_{(u,v)} IP_v} \quad (5)$$

Where $R_{(u,o)}$ is the predicted rating of location o to user u, $Sim_{(u,v)}$ is the similarity between user u and user *v*, IP_v is the value of social impact on user *v*, $Freq_{(v,o)}$ is the number of times a user *v* checks in at a place *o*.

4. EXPERIMENTS

To evaluate, our proposed method we compared it with two current location the recommendation methods. The First method combine CF technique and CBF together (called PTP) and the second method combined CF technique and social filtering together (called RSGI). All these method are apply on the same data sets [9,10] called the Gowalla dataset, which consists of the check-ins histories of each user, the friendships of each user. The experimental results can be used to measure performance.

4.1 Dataset

The collected data that used in this experiment was obtained from web-site named Gowalla, recorded in 12th December 2012. It consisted of; 407,533 users' data, 36,001,959 checked-in records, 4,418,339 friendship data and 15,209 location records and data. The given location data was randomly selected from 50 users, whom had checked-in record from 1 to 621 times in maximum to be able to create covered recommendation. By the cleaning of these data, since the need of the perfection of each user data appropriately, it was required for using these data as an initial data in these all three methods.

4.2 Evaluation metrics

In this work we use two evaluation metrices. They are coverage and NDCG.

4.2.1 Coverage

Coverage is an indicator of recommender system's coverage recommendation percentage. It shows how much the method can generate recommendation, which obtained by the calculation from equation 6.

$$Coverage = \frac{Number\ of\ predictable\ suggestions}{Number\ of\ all\ suggestion} * 100 \quad (6)$$

4.2.2 NDCG Score Average

NDCG is an indicator of the recommendation's efficiency that recommender system can provide to the users [11]. This is the evaluation of recommendation raking accuracy. The evaluation can be calculated by the calculation of equation 7.

$$NDCG_p = \frac{DCG_p}{IDCG_p} \quad (7)$$

Where $DCGp$ is a cumulative relevance score for all 5 positions (for this work). p is the position, which rank is 1 to 5 (for this work). $IDCG_p$ is an ideal cumulative relevance score for all 5 positions (for this work). This work try to measure accuracy of 5 position ranked list of each method.

4.3 Experimental

In the result comparison between the proposed method and other two methods, they must completely predicted top 5 ranking of every user and every methods to compared efficiency of result among these methods. The results was revealed as following below.

4.3.1 Coverage

Based on the Coverage calculation, provide the same the performance of the proposed method and the PTP method are 61.2% effective, while the RSGI method is the least effective at 48.2% (see Figure 1).

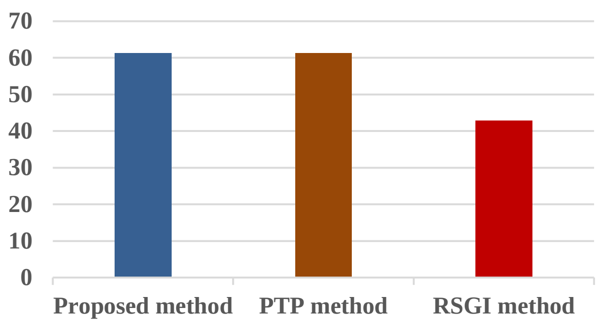

Figure 1 presents the coverage comparison among all three methods.

4.3.2 NDCG Average Score

When considering on the results of the calculation of the average efficiency, they reveal that the proposed method can provide the highest possible accuracy by the average at 90%. Furthermore, the PTP method and the NDCG average at 88% and the lowest average NDCG algorithm, is the RSGI method, with a score of about 83% (see Figure 2). The NDCG average score of all three methods is compared. The mean results are significantly different. Specifically, the results of the performance comparison between the proposed method and the RSGI method are up to 7%.

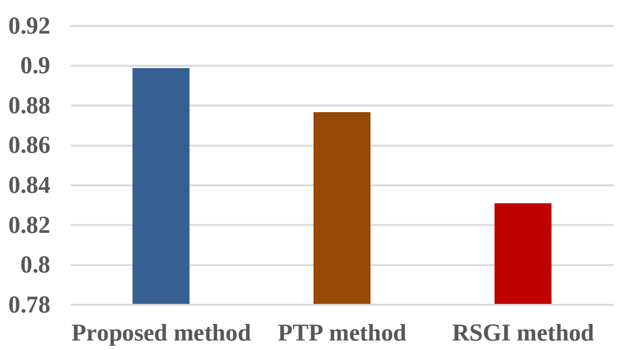

Figure 2 presents the NDCG Average Score comparison among all three methods.

5. DISCUSSION

5.1 Coverage

When compared the experimental result between proposed method and PTP method, it revealed that both method can create the same result, because both two methods had similar neighbor searching process, therefore it provided similar neighbor group. By using the same group of neighbors to assist the recommendation prediction to the users, the coverage of recommendation had the same result.

In addition, when compared the efficiency between the proposed method and RSGI method, it revealed that the coverage comparison result, the proposed method can provide a better result.. Since in the RSGI method, it used social influence to create the predication result. The social influence was created from both two users who had same number of online social friends. Therefore if both two users did not have same online social friend, it caused to unable to seek the social influence, thus it cannot calculate for rating prediction. Therefore the RSGI method cannot provide recommendation to the users who did not have same friend with someone in online social. In contrast, the proposed method used the social impact searching by using the number of followers normalized by the highest number of all user followers which it caused a capability to search for social impact for every user. As the result, the RSGI method cannot predict the recommendation to some of users, while the proposed method still can provide the predicted recommendation.

5.2 NDCG Average Score

From the experimental result, it revealed that the proposed method had the accuracy efficiency to provide the better quality recommendation to the users than other two methods.

When compared the efficiency between the proposed method and PTP method, it revealed that NDCG Average Score comparison, the proposed method can function better, even there was same number of neighbors and same checked-in records, but since the PTP method used only CF and CBF techniques for recommendation prediction. As you can see that, this method, it did not focus on neighbor weight in the on-line society. That means everyone have the same importance. In another hand, the proposed method was added the social filtering technique in the system, which focused on online social neighbor, but each neighbor had unequally importance. Therefore, it caused better accuracy efficiency of location recommendation to the users.

In addition, when compared the efficiency between the proposed method and RSGI method, it revealed that the proposed method can function better, since the RSGI method used only the CF and social filtering technique in providing the recommendation to the users which it was different from the proposed method. The proposed method also added CBF technique in recommendation providing which it causes more accuracy on searching the satisfied location for the users, focused on personal preference searching, while the RSGI method used only CF technique which focused on neighbor instead. Moreover, the using of social filtering technique in both two methods, it had different social influence/impact searching method. By the RSGI method, it used social influence searching method by having the same friends in on-line society, which affected especially in their group only. It provided uncovered social impact searching for everyone, in simply, it can provide only for some groups of users, but in turn, the proposed method used the social impact searching method, by using the number of followers, it can provide a better covered users' social impact method. In conclusion the proposed method had a better NDCG Average Score.

6. CONCLUSION

In this paper, we propose a new method of predicting the location to individual user by using three techniques: CF techniques, CBF techniques and social filtering techniques. It provides better efficient when compared to two current location recommendation methods. The first one uses CF and CBF together, and the other one uses CF and social filtering together.

The evaluation results show that the proposed method provide high accuracy than those two methods by using the NDCG measurement. Moreover, the coverage result provided a better result also. For further development, the system should record checked-in data with an additional information to create a covered information. It can solve the problem in predicting the recommendation for the user who had less or none checked-in record.

7. ACKNOWLEDGMENTS

This paper is supported the dataset from Lin *et al.* for research purpose.

8. REFERENCES

[1] Francesco Ricci, Lior Rokach, Bracha Shiapira, Paul B. Kantor "Recommender systems Handbook".

[2] Aggarwal, Charu C. "Recommender systems"

[3] John S. Breese, David Heckerman, and Carl Kadie "Empirical Analysis of Predictive Algorithms for Collaborative Filtering" in Technical Report MSR-TR-98-12, May 1998, revised October 1998.

[4] Long Guo, Jie Shao, Kian-Lee Tan, and Yang Yang "WhereToGo: Personalized Travel Recommendation for Individuals and Groups" in 2014 IEEE 15th International Conference on Mobile Data Management.

[5] Kyo-Joong Oh, Zaemyung Kim, Hyungrai Oh, Chae-Gyun Lim, and Gahgene Gweon "Travel Intention-Based Attraction Network for Recommending Travel Destination" in 2016 International Conference on Big Data and Smart Computing (BigComp).

[6] Shini Rengith, and Anjali C "A Personalized Travel Recommender Model Based on Content-based Prediction and Collaborative Recommendation" in International Journal of computer Science and Mobile Computing, ICMIC13, December-2013, pg. 66-73.

[7] Zhiwen Yu, Huang Xu, Zhe Yang, and Bin Guo "Personalized Travel Package with Multi-Point-of-Interest Recommendation Based on Crowdsourced User Footprints" in IEEE Transactions on Human-Machine System, Vol. 46, No. 1, February 2016.

[8] Da-Chuan Zhang, Mei Li, and Chang-Dong Wang "Point of Interest Recommendation with Social and Geographical Influence" in 2016 IEEE International Conference on Big Data (Big Data).

[9] Yong Liu, Wei Wei, Aixin Sun, Chunyan Miao, "Exploiting Geographical Neighborhood Characteristics for Location Recommendation" in Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM'14), pp. 739-748. ACM, 2014.

[10] Xin Liu, Yong Liu, Karl Aberer, Chunyan Miao, "Personalized Point-of-Interest Recommendation by Mining Users' Preference Transition" in Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM'13), pp. 733-738. ACM, 2013.

[11] Y. Wang, L. Wang, Y. Li, D. He, W. Chen & T.-Y. Liu. "A Theoretical Analysis of NDCG Ranking Measures" in JMLR: Workshop and Conference Proceedings vol. (2013)