Tourism Recommender System Based on Cognitive Similarity Between Cross-Cultural Users

Luong Vuong Nguyen, Tri-Hai Nguyen and Jason J. Jung [1]

Department of Computer Engineering, Chung-Ang University, Seoul 156-756, Republic of Korea

Abstract. Nowadays, the speedy increasing information in tourism services since a massive amount of data is constructed by tourists experiences. The recommendation systems are widely applied to tourism services and focus on determining personalized user preferences to handle this extensive information. Exploiting the different cultural effects rarely consider in recent studies despite this factor influences recommendation based on user preferences. Furthermore, existing research only evaluates the relevance of cultural differences to their recommendation, rather than using the cross-cultural factors to recommendations systems. This paper proposes the collaborative filtering recommendation system based on similar tourist places where users from different cross-cultural can share their spatial experiences. To do that, we first collect user feedback about similar tourist places from many nationalities (consider as the cultures). We then exploit this feedback to define similar cross-cultural users (neighbors) based on a cognitive similarity. Finally, the system generates personalized recommendations based on user experiences and their neighbors. The initial dataset collected from TripAdvisor, consisting of four types such as hotels, restaurants, shopping malls, and attractions, is provided to the feedback collection function in our experiment. We were using the classical method, user-based Pearson correlation, as a baseline to demonstrate the performance of our proposed method. The result shows that the proposed system outperforms the baseline in terms of MAE and RMSE metrics.

Keywords. recommendation system, cross-cultural, crowdsourcing, cognitive similarity

1. Introduction

Today, information about travel places and relevant resources, such as accommodations, restaurants, shopping malls, is usually searched by tourists who plan a trip. However, this information increases very fast every day, leading to a very complicated and time-consuming evaluation. Many systems are created to handle and make suitable recommendations depending on tourist preferences to solve this problem. In recent researches, the personalized techniques applied to collaborative filtering recommendation systems (RS) are commonly used in the tourism domain [1,3,4]. In proposing schemes, collabo-

---

[1] Corresponding Author: Professor, Chung-Ang University, Seoul 156-756, Republic of Korea; E-mail: j2jung@gmail.com.

rative filtering (CF) approaches are the most used. They make use of user-provided details related to assessments (or ratings). This can cause the issue of sparsity when user assessments are inadequate. Traditional CF approaches based on nearest neighbor algorithms, on the other hand, show significant issues with efficiency and scalability. In the last few years, several recommendation strategies have been suggested to improve the efficiency of the recommendations [2,5] and address other typical disadvantages of the recommended systems.

In this study, we propose a novel approach that determines the cognitive similarity between cross-cultural users to improve the performance of CF effectively. To do it, we first collect feedback about the tourist places from users who have visited or know these places. This feedback content includes what they think about the tourist places (e.g., they like it, and they think it similars others). Based on this feedback, the second step is to construct the cognition pattern of each user that is the priority of them with the extracted factor from tourist places. These extracted factors from tourist places are the area ranking, the user rating, and the number of reviews. The cognitive similarity between the users (the similar cognition pattern) is estimated using cosine similarity measurement. Finally, we make the recommendations to active users according to the k-nearest neighbor algorithm. The initial dataset for the crowdsourcing platform is collected from TripAdvisor consists of four types of tourist places: hotels, restaurants, shopping malls, and attractions. All feedback from users about tourist places stored in the database and provide for the experiments. The main contributions of this paper are summarized as follows.

1. We deploy the online crowdsourcing platform, where users from cross-cultural can share and give their feedback about tourist places that they have visited or known.
2. We do extensive experiments that are modeling the cognitive similarity-based CF recommendation system of cross-cultural users for tourism services

The remainder of this paper is structured as follows. Section 2 introduces the related work on the CF recommendation systems and recent studies in improving the user-based method. Section 3 presents the tourism recommendation system based on the cognitive similarity between cross-cultural users. The experimental results are details describe in Section 4. In Section 5, we provide some conclusions and directions for future work.

2. Related work

RS techniques, such as content-based, collaborative filtering, demographic filtering, have been applied to tourism services to provide personalized recommendations to tourists. Among these techniques, the CF approach was widely applied in the tourism area. In general, CF can be classified into two classes model-based and neighborhood-based methods. Model-based methods make recommendations by learning a predictive model according to user-item ratings. There are several recent studies in model-based, such as latent semantic analysis (LSA), Bayesian clustering, support vector machines (SVM), latent Dirichlet allocation (LDA), and singular value decomposition (SVD) [7,8,9,10]. Neighborhood-based methods make recommendations by measuring the similarities among users or among items (know as user-based and item-based, respectively). In recent research, the effectiveness of CF stability-enhancing was mentioned in [13] that focused

on clustering users method. In particular, they presented a novel bio-inspired clustering based on swarm intelligence and fuzzy clustering models. Besides, in [14], ontology and dimensionality reduction techniques were used to handle both popular RS problems: sparsity and scalability. This research aims to increase the accuracy in the recommending process by using the ontology, combine with Singular Value Decomposition (SVD), the dimensionality reduction technique, to improve the scalability of recommending method. In [15], the cognitive similarity between users was considered to define the similar users that improve the performance of collaborative filtering in the movie recommendation system. This issue was also mentioned and exploited in [11] that proposed the crowd-sourcing platform to collecting feedback from users who have experiences in the movies domain. In [12], the cross-cultural contextualization was exploited for the recommendation system. This study proposed a new definition of cross-cultural contextualization, then based on that, they compute the cross-cultural factor affecting users. To do that, they presented a model based on the combination of matrix factorization and clustering techniques.

This work focuses on improving a user-based recommendation system that proposes a better method to define the most similar items for active users. We define the k-nearest neighbor dynamically based on cognitive similarity among users in each cluster of users and the whole users in the system because the cognitive similarity between users dynamically changes according to their history activities. In this way, the proposed method improves the accuracy in define similar users, and therefore, the accuracy in the recommendation process is enhancing. The evaluation demonstrates that the proposed method achieves the outperformance in comparison with the traditional method, user-based user-based Pearson correlation.

3. Recommendations based on Cognitive Similarity between Cross-Cultural Users

3.1. Similarity Measurement

In our dataset, there are four types of tourist places, and therefore we decided to use the overlap extracted factors from these tourist places, which are (R) user ratings, (K) area rankings, and (N) the number of reviews. To calculate the similarity between tourist place p_i and p_j, we following these steps *(i)* each tourist place p represented by these overlap extracted factors, *(ii)* determine the similarity of each pair of extracted factors in the first step by applying the cosine similarity formulation, *(iii)* the average similarity score of all extracted factors is assigned to a similarity score between tourist place p_i and p_j. The formulation for measuring the similarity score ($\Im$) between tourist place p_i and p_j is described by

$$\Im(p_i, p_j) \equiv \langle R_{ij}, K_{ij}, N_{ij} \rangle, \tag{1}$$

where R_{ij}, K_{ij}, and N_{ij} represent the similarity measurement of these factors user ratings, area rankings, and the number of reviews between tourist place p_i and p_j, respectively. In this study, we measure the similarity between extracted factors by using the soft cosine metric defined as the cosine of the angle between two non-zero vectors of an inner product space. For example, considering the users rating factor, the similarity of these

extracted factors between tourist places p_i and p_j measured by soft cosine metric, which described as

$$R_{ij} = \frac{\sum_1^n R_i R_j}{\sqrt{\sum_1^n (R_i)^2} \sqrt{\sum_1^n (R_j)^2}}. \tag{2}$$

We repeat Eq. 2 with the remaining factors that are area ranking (K_{ij}) and the number of reviews (N_{ij}). Finally, we obtain the ($\Im$) between these tourist places according to Eq. 1. According to these steps above, the formulation Eq. 1 is repeated to measuring similarity between p_i and each tourist place in the whole of our dataset. We then obtain a set $\{\Im(p_i, p_{i+j}) | j \in [1,..,n-1]\}$, where n is number of the tourism places in the database.

3.2. Cognitive Similarity-based Cross-Cultural Users Clustering

We collect feedback from cross-cultural users that they think about the tourist places (e.g., similar, interest) based on the crowdsourcing platform using the initial dataset. The process to collect feedback is as follows.

1. The system randomly shows four tourist places (consist of hotels, restaurants, shopping malls, and attractions). The user selects the tourist place that he/she has visited or known (implicit that he/she was interested in this place).
2. The platform will suggest four tourist places based on the type of tourist places that he/she selected and the list $N - places$ constructed as described in Section 3.1.
3. He/she selects one of four tourist places that they think similar and submit, the system store this information and assigned it to the history of user activities (pair of similar places). If he/she does not select any tourist place, the system assigns the tourist place that he/she chose in the first step to the history of user activities (interest tourist place).

Following this process, we have datasets of each user in the system. The essential property in these datasets is the priority model measured during selecting similar tourist places' process from each user (called the user cognition pattern). This cognition pattern is updated according to the activities of users in the platform. We then cluster the users based on these similar cognition patterns of the cross-cultural users. In this study, the similarity between cognition patterns, which we called cognitive similarity between cross-cultural users, is determined.

Definition 1 - Cognitive Similarity between Cross-Cultural Users.

Given user u and v, the cognitive similarity $\Im$ between these users is the similarity between their cognition pattern that represents by the similarity of their priorities in selecting similar tourist places. By using the cosine similarity, we have the following formula

$$\Im(u,v) = \frac{\sum_i^N \rho_u^i \rho_v^i}{\sqrt{\sum_i^N (\rho_u^i)^2} \sqrt{\sum_i^N (\rho_v^i)^2}}, \tag{3}$$

where ρ represent the priorities in selecting similar tourist places of user u and v.

For instance, given i is the number of tourist places in the datasets $I(u)$ of user u, the cognition pattern ρ_u^i of user u is defined as a set $\{R, K, N\}$ which each item in a set have value is the average similarity score of each extracted factor from each pair of similar tourist places in $I(u)$. When user u has new activities (new data inserted to $I(u)$), the cognition pattern of user u automatically re-calculated and updated to the database. A set $\{R, K, N\}$ has three value scores re-ordered from highest to smallest and is an ordered vector, which is described as

$$\rho_u^i = \{R_u^i; K_u^i; N_u^i\}. \tag{4}$$

In this work, we aim to make the user clustering automatically, and dynamic re-calculate for updating to the database since any new activities of the user will make the cognition pattern changing. We deploy the calculation function to determine the cognition pattern of each user according to any new activities that happened in the whole platform. It means on the servers side, and the platform has to manage the database structure to store and handle the long complicate process.

4. Experiment

4.1. Dataset

To construct the initial datasets provided for the crowdsourcing platform, we decide to crawl tourist places information from TripAdvisor, which has the practical information for our purpose. We have collected approximately eighteen thousand tourist places from 84 cities of 24 countries distributed on five continents. The detailed statistic of the initial tourist places dataset is shown in table 1.

We keep collecting data from more than 150 active users on our platform and enrich our database. Our purpose is to deploy the system scalability to as many users as possible for collecting more valuable feedback from many different users from different countries (different cultures). The collected feedback from cross-cultural users have the format $(U_i, p_a, p_b, \rho_{U_i}^{a,b}, \gamma)$ inside U_i is the id of the user. p_a and p_b are id of tourist places, respectively. $\rho_{U_i}^{a,b}$ is a vector representing the cognition pattern of user U_i in selecting a pair similar tourist places (p_a, p_b).

Table 1. Statistics of the tourist places dataset

#	Title	Amount
1	Number of countries / cities / tourist places	24 / 84 / 18,286
2	Number of reviews	239,446
3	Number of ratings	166,428
4	Number of users	153
5	Number of feedback from users	2210

4.2. Evaluation

To construct the evaluation report, we experimented on the dataset described above. The steps to deploy the experiment is described as follows. Firstly, the dataset is separated into five mutually exclusive folds because we decide to carry out the evaluation, which uses the five-fold cross-validation. We split the dataset into two parts which 80% for the training-set and the rest 20% for the testing-set each fold. In the second step, the tourist places prediction in the testing-set according to the corresponding tourist places in the training-set for each fold. Finally, the performance is evaluated based on comparing the predictions and the authenticity available in the test-set for each fold. The final result is an average of all predicted in all folds.

The classic method, User-based Pearson Similarity (UBPS), is identified as the basis for the comparative analysis in this section. Among the different criteria used to evaluate the accuracy of CF approaches, the mean absolute error (MAE) and the root mean square error (RMSE) is selected as the evaluation metric for the comparative analysis report. The formulations of MAE and RMSE are described by

$$\textbf{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - y_i^p|, \tag{5}$$

$$\textbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - y_i^p)^2}, \tag{6}$$

where n is the number of tourist places, y_i is the real values and y_i^p is the predicted values in the test set. The results of MAE and RMSE range from 0 to infinity. Infinity is the maximum error according to the scale of the measured values. The baseline has the best performance in approximately 50 neighbors of the active user. Therefore, we designate the experiments with the neighborhood sizes to be $\{5, 10, 20, 30, 50\}$. The comparison between the proposed method and the baseline is shown in Figure 1.
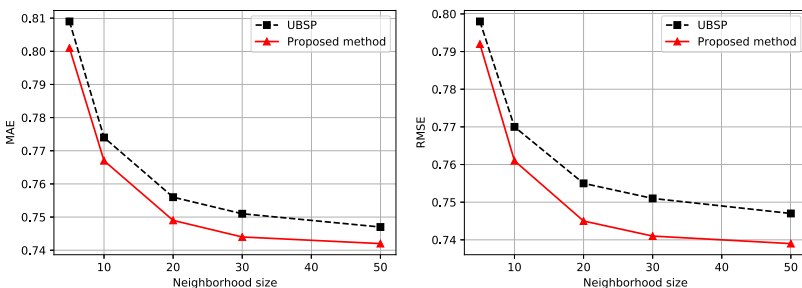

Figure 1. Performance comparison between the proposed method and the baseline in term of MAE and RMSE

5. Conclusion and Future work

This paper presented a novel approach to improving user-based CF that focuses on user clustering based on the cognitive similarity between cross-cultural users. We deployed a crowdsourcing platform to gather feedback from cross-cultural users for this purpose. The platform has the simplest and easiest steps in the feedback collection process from users. The experiments performed on the data set verified that the proposed approach meets the expectations that improve performance over the baseline UBPS, which has only global similarity.

In the future study, the valuation methods used to form a user cluster shall be examined based on data obtained from cross-cultural users. For example, the method will consider the cognition pattern of users and consider the demographic or other personal information that is allowed to collect from users. Moreover, a more advanced and sophisticated approach to splitting the list into the preferred user cluster can be introduced.

References

[1] Chen R, Hua Q, Chang YS, Wang B, Zhang L, Kong X. A Survey of Collaborative Filtering-based Recommender Systems: From Traditional Methods to Hybrid Methods based on Social Networks. IEEE Access. 2018 Oct 24;6:64301-20.

[2] Haruna K, Akmar Ismail M, Suhendroyono S, Damiasih D, Pierewan AC, Chiroma H, Herawan T. Context-Aware Recommender System: A Review of Recent Developmental Process and Future Research Direction. Applied Sciences. 2017; 7(12):1211.

[3] Wang X. Personalized Recommendation Framework Design for Online Tourism: Know You Better Than Yourself. Industrial Management & Data Systems. 2020 Oct; 120(11):2067-2079.

[4] Nguyen LV, Jung JJ, Hwang M. OurPlaces: Cross-Cultural Crowdsourcing Platform for Location Recommendation Services. ISPRS International Journal of Geo-Information. 2020 Nov; 9(12):711.

[5] Nguyen LV, Nguyen TH, Jung JJ. Content-Based Collaborative Filtering using Word Embedding: A Case Study on Movie Recommendation. Proceedings of the International Conference on Research in Adaptive and Convergent Systems (RACS '20); 2020 Oct 13-16; Gwangju, South Korea. ACM; 2020. p. 96—100.

[6] Herlocker JL, Konstan JA, Borchers A, Riedl J. An Algorithmic Framework for Performing Collaborative Filtering. Proceedings of SIGIR Forum; 2017 Aug 2; New York, USA. ACM; 2017. p. 227–234.

[7] Hofmann T. Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis. Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval; 2003 Aug 1; Toronto, Canada. ACM p.259–266

[8] Breese JS, Heckerman D, Kadie C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. arXiv preprint arXiv:1301.7363. 2013 Jan 30.

[9] Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. Journal of Machine Learning Research. 2003 Jan 3:993-1022.

[10] Adomavicius G, Kwon Y. New Recommendation Techniques for Multicriteria Rating Systems. IEEE Intelligent Systems. 2007 Jun; 22(3):48-55.

[11] Nguyen LV, Jung JJ. Crowdsourcing Platform for Collecting Cognitive Feedbacks from Users: A Case Study on Movie Recommender System. In: Reliability and Statistical Computing. Springer; 2020. p.139-150.

[12] Hong M, An S, Akerkar R, Camacho D, Jung JJ. Cross-cultural Contextualisation for Recommender Systems. Journal of Ambient Intelligence and Humanized Computing. 2019 Sep;9:1-2.

[13] Logesh R, Subramaniyaswamy V, Malathi D, Sivaramakrishnan N, Vijayakumar V. Enhancing Recommendation Stability of Collaborative Filtering Recommender System through Bio-inspired Clustering Ensemble Method. Neural Computing and Applications. 2020 Apr;32(7):2141-64.

[14]   Nilashi M, Ibrahim O, Bagherifard K. A Recommender System based on Collaborative Filtering us-
       ing Ontology and Dimensionality Reduction Techniques. Expert Systems with Applications. 2018 Feb
       1;92:507-20.
[15]   Nguyen LV, Hong MS, Jung JJ, Sohn BS. Cognitive Similarity-based Collaborative Filtering Recom-
       mendation System. Applied Sciences. 2020 Jan;10(12):4183.