

## Association for Information Systems AIS Electronic Library (AISeL)

---

PACIS 2016 Proceedings

Pacific Asia Conference on Information Systems  
(PACIS)

---

Summer 6-27-2016

# HOTEL RECOMMENDATION SYSTEM BASED ON REVIEW AND CONTEXT INFORMATION: A COLLABORATIVE FILTERING APPRO

Ya-Han Hu

*National Chung Cheng University, yahan.hu@mis.ccu.edu.tw*

Pei-Ju Lee

*National Chung Cheng University, pjlee@mis.ccu.edu.t*

Kuanchin Chen

*Western Michigan University, kc.chen@wmich.edu*

J. Michael Tarn

*Western Michigan University, mike.tarn@wmich.edu*

Duyen-Vi Dang

*, National Chung Cheng University, dangduyenvi@gmail.com*

Follow this and additional works at: <http://aisel.aisnet.org/pacis2016>

---

### Recommended Citation

Hu, Ya-Han; Lee, Pei-Ju; Chen, Kuanchin; Tarn, J. Michael; and Dang, Duyen-Vi, "HOTEL RECOMMENDATION SYSTEM BASED ON REVIEW AND CONTEXT INFORMATION: A COLLABORATIVE FILTERING APPRO" (2016). *PACIS 2016 Proceedings*. 221.

<http://aisel.aisnet.org/pacis2016/221>

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2016 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# HOTEL RECOMMENDATION SYSTEM BASED ON REVIEW AND CONTEXT INFORMATION: A COLLABORATIVE FILTERING APPROACH

Ya-Han Hu, Department of Information Management, National Chung Cheng University,  
Chiayi, Taiwan, yahan.hu@mis.ccu.edu.tw

Pei-Ju Lee, Department of Information Management, National Chung Cheng University,  
Chiayi, Taiwan, pjlee@mis.ccu.edu.tw

Kuanchin Chen, Department of Business Information Systems, Western Michigan  
University, Kalamazoo, Michigan, 49008-5412, USA, kc.chen@wmich.edu

J. Michael Tarn, Department of Business Information Systems, Western Michigan  
University, Kalamazoo, Michigan, 49008-5412, USA, mike.tarn@wmich.edu

Duyen-Vi Dang, Department of Information Management, National Chung Cheng  
University, Chiayi, Taiwan, dangduyenvi@gmail.com

## Abstract

*Due to the increment of different formats of online expressions such as reviews, ratings, and recommendation, it is getting more difficult to identify users' preferences toward the products. A large number of reviews can be generated and diffused by online users in travel booking websites. A set of Recommendation Systems (RSs) has emerged to help consumers to filter items based on their preferences. The Collaborative Filtering (CF) based approach is one of the most popular techniques of the RS; however, it also suffers from several fundamental problems such as data sparsity, cold start, shortage, and rating bias. This study proposes a context-aware hotel recommendation (CAPH) approach; using the context-aware information to provide personalized hotel recommendation system. This research considers recommending hotels based on the hotel features and traveler's type. Experimental data is collected from Tripadvisor.com during the period of 2015 to 2016. The evaluations of system accuracy will be conducted and then compared with the user-based / item-based CF model.*

*Keywords: Recommender system, context analysis, collaborative filtering.*

# 1 INTRODUCTION

Given the rapid development of Web 2.0, the amount of online content information has increased dramatically, which caused the problem of information overload in big data era (Abbas, Zhang, & Khan, 2015). It is difficult to capture end-users' preferences toward features of an item from various different forms of online expression such as reviews, ratings, and recommendations. A personalized recommendation system (RS) (Adomavicius & Tuzhilin, 2011) is one effective way to help customers filtering our information. The Collaborative Filtering (CF) based approach is the most successful technology among RSs (Sun, Wang, Cheng, & Fu, 2014) and it can be classified as User-based CF (UCF) model or Item-based CF (ICF) model. The UCF model provides recommendations by computing similar neighbor and create a group of like-minded users with a target user (Jin & Chen, 2012). The ICF model computes similarity based on items, which find similar items to the given user's rated items (Pappas & Popescu-Belis, 2013). Many researchers (Kim, Alkhalidi, El Saddik, & Jo, 2011; Pappas & Popescu-Belis, 2015) have studied on the effectiveness of CF applies on e-commerce.

The CF suffers from several fundamental problems such as data sparsity, cold start, shortage, and rating bias; and the data sparsity problem is the most important one (Moshfeghi, Piwowarski, & Jose, 2011). To address this issue, some researchers have explored the rich-user-items-information approach (Liu, He, Wang, Song, & Du, 2013) to fix it.

According to Travel Statistics for Tour Operators<sup>1</sup>, there are more than 148.3 million people make reservations for their accommodations, tours, and activities through the internet, which is more than 57% of all travel reservations every year. Furthermore, the role of Cyber Travel Agents (CTAs) such as TripAdvisor.com, booking.com, Venere.com, etc. dramatically influence the tourism landscape and hospitality phenomena (Marchiori, Eynard, Inversini, Cantoni, & Cerretti, 2011). The effective use of CF techniques in CTAs can increase the possibility of online booking and buying.

The user context information such as traveler types (e.g. family or business) is an important factor for RSs on recommending hotels. To generate recommendations for travellers is inherently difficult because of the involvement of experience goods is high and the hotel quality is often unknown before consumption (Forman, Ghose, & Wiesenfeld, 2008). Recently, the context-aware recommendation system (CARS) is a popular technique to deal with information filtering problem (Wachsmuth, Trenkmann, Stein, & Engels, 2014). CARS explores customers interests and presents information on items to match their preferences using context information (Adomavicius & Tuzhilin, 2011). Although CARS has been applied to hotel recommendation in many applications, the contextual modeling approach also suffers from the data sparsity problem.

To tackle the challenges above, this study develops a context-aware personalized hotel (CAPH) recommendation system. First, users' preferences for hotels will be filtered from user online reviews (Liu et al., 2013), then the representative terms in the reviews can be considered as inferred preference ratings and can be corporate into the user-feature matrix in the imputation technique ; therefore, the matrix in cases of imputation is built by representative preference results to generate the neighbourhood-based CF models. Next, the denser matrix from UCF (Jin & Chen, 2012) must be integrated into inferred rating data. The experimental data is collected from TripAdvisor.com dataset and a set of evaluations about the prediction accuracy of rating will be conducted.

---

<sup>1</sup> <https://www.rezdy.com/resource/travel-statistics-for-tour-operators/>

## 2 LITERATURE REVIEW

The CF has been explored in many domains of social media (Sun et al., 2014), restaurant (Liu et al., 2013), and travel (Zheng, Burke, & Mobasher, 2012) with the objective of making recommendations to consumers. The similar users or items are identified using a similarity metric (Pappas & Popescu-Belis, 2015) and the evaluation techniques include Pearson Correlation Coefficient, Spearman Rank Correlation Coefficient, Cosine Similarity, and Mean-Square Difference. A subset of similar users or items will be determined by the most similar neighbours of a target user or item. There are two methods usually used for neighbourhood selection: the similarity thresholding (Gao & Li, 2010) and the top K-nearest neighbor (Sun et al., 2014). The preference prediction researches focus on Recommendation task (Kim et al., 2011) and the Rating prediction task (Sun et al., 2014). In the real world (e.g. social media site or e-commerce site), each individual user has expressed their preferences only on an extremely small portion of the products. Therefore, the system is generally insufficient for identifying similar neighbours for lacking intersection between users or items (Sun et al., 2014). (Hu, Dou, & Liu, 2012) demonstrated that the CARS can explore consumers' interests and present information on items that match consumers' preferences based on context information. (Adomavicius & Tuzhilin, 2011) showed there are three recognized paradigms that the contextual information incorporated into RS are contextual pre-filtering, contextual post-filtering, and contextual modeling. In this study, the contextual pre-filtering approach is selected for its straightforward processing and flexibility of justification so the CF can be utilized before or after computing predictions.

## 3 RESEARCH METHODOLOGY

The proposed CAPH system is shown in Figure 1. This work intends to crawl ratings and reviews based on top contributor users' profiles (i.e., the users having written more than 20 hotel reviews in TripAdvisor.com). The types of travellers are treated as contextual information and assigned to each rating. The given opinions of the consumers reflect particular traveller's type (i.e., solo or friends).

### 3.1 Review pre-processing

This paper uses Google Spell Check to correct the grammatical errors and typos in the crawled documents. Second, the Stanford CoreNLP toolkit (Manning et al., 2014) is used to categorize words with similar grammatical properties through Part of speech tagging system (POS). The third step is to composite weights of each term in the review using Term Frequency-Inverse Document Frequency (TF-IDF) technique to determine which words in a corpus of documents might be the most representative (Pappas & Popescu-Belis, 2015). The fourth step is that hotel features are defined and the representative terms are assigned to the appropriate hotel features. Based on the predefined hotel features used in existing studies, our study identifies twelve types of hotel features of food, room, cleanliness, services, sleep quality, location, value, security, sport facilities, general amenities, entertainment, and prices that were likely to represent customers' consideration (Albaladejo-Pina & Díaz-Delfa, 2009). We use Normalized Google Distance (NGD) to measure the distance between representative terms and hotel features (Cilibrasi & Vitányi, 2007). The pre-filtering approach on CARS is performed in this study. The rating and review of each traveller's type reflect the reviewer's opinion toward a particular traveller's types. For this context, the subset of the traveller's type includes Families, Couples, Solo and Business; so reviews and ratings for an item will be split into four subsets according to the value of a contextual variable.

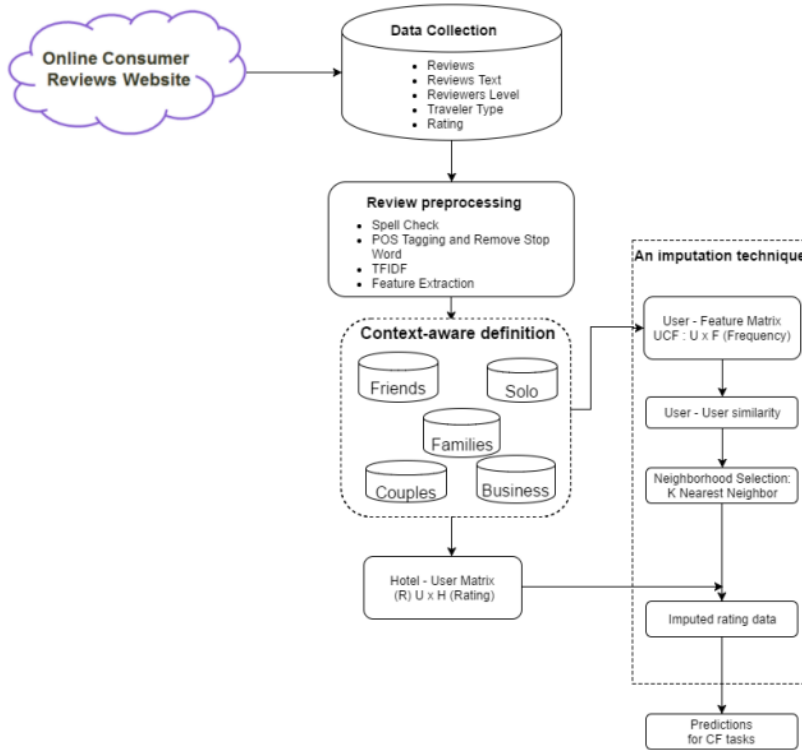


Figure 1. CAPH system

### 3.2 Rating imputations in CAPH

The first step is to build the user's profile from the features which users consider about items over time. We use the user-feature matrix to predict a user's preference in a specific hotel based on the feature information from the similar user profile. Figure 2 shows a user-feature matrix A. There is a set of users  $U = \{u_1, u_2, \dots, u_m\}$  and a set of hotel features  $F = \{f_1, f_2, \dots, f_j\}$ . Each user  $u_a$  has a list of features  $f_k$  and each element  $N(u_a, f_k)$  indicates that the number of hotels considered by user  $u_a$  with the specific feature  $f_k$ . The hotel features are used as the representative items as well as the basis of the similarity computation. The cosine similarity between  $u_a$  and  $u_v$  (Adomavicius & Kwon, 2007) is computed based on the matrix A.  $N(u_a, f_k)$  and  $N(u_v, f_k)$  denote the feature frequencies of users  $u$  and  $v$ . After computing similarity of all user pairs, we get the preference similarity matrix and the similarity score is in the range of [0,1].

A	$f_1$	$f_2$	...	$f_j$
$u_1$	1	2	...	4
...				
$u_a$	4	3	...	$N(u_a, f_k)$
...				
$U_m$	4	5	...	5

Figure 2. User - Feature Matrix

B	$f_1$	$f_2$	...	$f_j$
$h_1$			...	4
...				
$h_i$	4		...	$S(h_i, f_k)$
...				
$h_n$		1	...	

Figure 3. The sparse Hotel - Feature Matrix

In CAPH system, the rating imputation is based on the user-hotel matrix  $R$  (as shown in Figure 3) and the user-feature matrix  $A$ 's similarity's weight. The expected rating that a user  $u_a$  would give to a hotel  $h_i$  is marked as  $R'(u_a, h_i)$ . The rating score is based on the similarity weight from UCF and defined as:

$$R'_{ai(u)} = \bar{R}_a + \sum_v^k (R_{iv} - \bar{R}_v) \cdot \text{sim}(a, v) / \sum_v^k \text{sim}(a, v) \quad (1)$$

Where  $R'_{ai(u)}$  is the predicted rating for the given user  $a$  on an unrated item  $i$ ,  $\bar{R}_a$  and  $\bar{R}_v$  are the mean ratings of user  $a$  and  $v$  rate over all items,  $R_{iv}$  is the rating of user  $v \in k$  for item  $i$ .  $v \in k$  designates the top  $k$  nearest neighbours of user  $a$ . According to user-feature matrix  $A$ , the  $\text{sim}(a, v)$  denotes the weight of similarity between user  $a$  and user  $v$ . Finally, we could get the denser hotel-user matrix  $R'$  (Figure 3) to make hotel recommendation on CF framework and generate two corresponding enhanced models: CAPH\_UCF and CAPH\_ICF.

This research considers every user has his/her consideration for different hotel features. Based on running this system into a context-aware situation by predicting the expected rating under the specific traveller's type (e.g., business trip), it is possible to recommend items that would be appropriate to the user based on the hotel feature and traveller's type.

$R'$	$h_1$	$h_2$	...	$h_i$	...	$h_m$
$u_1$	1	2	...	4	...	5
-	-	-	-	-	-	-
$u_a$	4	3	...	$R(u_a, h_i)$	...	$h_i$
$U_n$	4	5	...	5	...	3

Figure 4. User- Hotel matrix

## 4 EXPERIMENTAL EVALUATION

The performance of the CAPH system was assessed using the leave-one-out cross-validation (all-but-one strategy): for a given user, taking turns to select one rated item, perform training on the remaining data, and then compare the rating outcomes with the recommended and the withheld elements.

### 4.1 Data

In order to evaluate the proposed CAPH framework, we focus on hotels in the USA from TripAdvisor.com. We crawled ratings and reviews from this website based on user's profile whose review badge is a senior contributor or top contributor (i.e., having written more than 20 reviews). We selected the traveller type as contextual information and assigned it to each rating given the opinion of the consumer reflected on a particular traveller type, which is: Friends and Solo (both have sparsity less than 0.69%). The original dataset contains 279 users, 8257 hotel, and 2092 ratings and reviews.

### 4.2 Experimental setup and performance measures

Our system architecture was implemented in R language and tested on Intel core i5-4570 3.2GHz Windows 8.1 system with 16 gigabytes of main memory. In experimental evaluation, we chose the typical contextual recommendation as the baseline. For the performance evaluation of CAPH system on a given data, we aim to predict the rating that the user would give to a target item;

therefore, the performance of the CAPH system is assessed by measuring the accuracy of rating predictions.

In RSs, the root mean squared error (RMSE) and the mean absolute error (MAE) are the most commonly used performance measures. The formulas of these measurements are given by equations (4) and (5), where  $T$  is the total number of ratings in the training data, the system generates predicted ratings  $\widehat{r}_{ui}$  for user  $u$  on item  $i$  and the actual ratings  $r_{ui}$  are known.

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{(u,i)} (\widehat{r}_{ui} - r_{ui})^2} \quad (4)$$

$$\text{MAE} = \frac{1}{T} \sum_{(u,i)} |\widehat{r}_{ui} - r_{ui}| \quad (5)$$

### 4.3 Results

According to table 1, the experimental results have demonstrated that the proposed method improves the baseline’s sparse matrices effectively. In the condition of baseline, because of highly sparse matrix problem the both values of UCF’s RMSE and MAE will be N/A when the neighbourhood size is 30. In our method, the values of CAPH\_UCF’s MAE and RMSE are easily obtained. Furthermore, when the matrices are covered about 35% of the original rating data, our method achieves higher in CAPH\_ICF’s MAE value (Friends: 0.435 and Solo: 0.234) while the baseline is 0.667 in Friends type and 1.25 in Solo type.

Travel type	Method		MAE	RMSE
Friends	Baseline	UCF	N/A	N/A
		ICF	0.667	0.745
	Our research	CAPH_UCF	0.816	1.040
		CAPH_ICF	0.436	0.567
Solo	Baseline	UCF	N/A	N/A
		ICF	1.25	1.323
	Our research	CAPH_UCF	0.271	0.297
		CAPH_ICF	0.235	0.326

Table 1. Results of the CAPH and the baseline method.

## 5 CONCLUSION

In this research, we proposed a CAPH recommender system for the use of CTAs. This study comprehended the researches of CF, Context-aware, hotel RS, and CARS with time constraints and developed the CAPH system based on UCF and ICF for rating prediction. The proposed method utilizes review text to represent rating data preference and makes recommendations from the denser matrix than the original sparse matrix. After a comparative analysis of the results, the findings indicate that users’ reviews can overcome the sparsity of typical contextual RS effectively. In the future work, we are going to complete the experiment with more data, modify the context, and perform evaluations of the experimental results.

## Acknowledgement

This research was supported in part by the Ministry of Science and Technology of the Republic of China (grant number MOST 104-2410-H-194-070-MY3)

## References

- Abbas, A., Zhang, L., & Khan, S. U. (2015). A survey on context-aware recommender systems based on computational intelligence techniques. *Computing*, 97(7), 667–690.
- Adomavicius, G., & Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. *IEEE Intelligent Systems*, 22(3), 48–55.
- Adomavicius, G., & Tuzhilin, A. (2011). Context-Aware Recommender Systems. In *Recommender Systems Handbook* (pp. 217–253). Boston, MA: Springer US.
- Albaladejo-Pina, I. P., & Díaz-Delfa, M. T. (2009). Tourist preferences for rural house stays: Evidence from discrete choice modelling in Spain. *Tourism Management*, 30(6), 805–811.
- Cilibrasi, R. L., & Vitányi, P. M. B. (2007). The Google Similarity Distance. *IEEE Transactions on Knowledge and Data Engineering*, 19(3), 370–383.
- Ding, X., Liu, B., Liu, B., Yu, P. S., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. *Proceedings of the International Conference on Web Search and Web Data Mining - WSDM '08*, 231.
- Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality. In *2013 46th Hawaii International Conference on System Sciences* (pp. 3119–3128). IEEE.
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research*, 19(3), 291–313.
- Gao, H., & Li, W. (2010). A hotel recommendation system based on collaborative filtering and Rankboost algorithm. *2010 International Conference on MultiMedia and Information Technology, MMIT 2010, 1*, 317–320.
- Jin, J., & Chen, Q. (2012). A trust-based Top-K recommender system using social tagging network. *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*.
- Kim, H.-N., Alkhaldi, A., El Saddik, A., & Jo, G.-S. (2011). Collaborative user modeling with user-generated tags for social recommender systems. *Expert Systems with Applications*.
- Li, G., Law, R., Vu, H. Q., Rong, J., & Zhao, X. (Roy). (2015). Identifying emerging hotel preferences using Emerging Pattern Mining technique. *Tourism Management*, 46, 311–321.
- Liu, H., He, J., Wang, T., Song, W., & Du, X. (2013). Combining user preferences and user opinions for accurate recommendation. *Electronic Commerce Research and Applications*.
- Manning, C. D., Bauer, J., Finkel, J., Bethard, S. J., Surdeanu, M., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55–60.
- Marchiori, E., Eynard, D., Inversini, A., Cantoni, L., & Cerretti, F. (2011). Harvesting Online Contents: an Analysis of Hotel Reviews Websites. In *Information and Communication Technologies in Tourism* (pp. 101–112). Springer Verlag.
- Moshfeghi, Y., Piwowarski, B., & Jose, J. M. (2011). Handling data sparsity in collaborative filtering using emotion and semantic based features.
- Pappas, N., & Popescu-Belis, A. (2015). Combining content with user preferences for non-fiction multimedia recommendation: a study on TED lectures. *Multimedia Tools and Applications*.
- Hu, R., Dou, W., & Liu, J. (2012). A context-aware collaborative filtering approach for service recommendation. In *2012 International Conference on Cloud and Service Computing* (pp. 148–155). IEEE.
- Sun, J., Wang, G., Cheng, X., & Fu, Y. (2014). Mining affective text to improve social media



item recommendation. *Information Processing & Management*.  
Zheng, Y., Burke, R., & Mobasher, B. (2012). Differential Context Relaxation for Context-Aware  
Travel Recommendation. In *Artificial Intelligence Review* (Vol. 35, pp. 88–99).