ENHANCEMENT OF USER PROFILING FOR TOURISM RECOMMENDATION SYSTEM

Pijitra Jomsri\textsuperscript{1}, Worasit Choochaiwattana\textsuperscript{2}

\textsuperscript{1}Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, Thailand
\textsuperscript{2}College of Creative Design and Entertainment Technology Dhurakij Pundit University, DPU Bangkok, Thailand

\textsuperscript{1}pijitra.jo@gmail.com, \textsuperscript{1}pijitra.jo@ssru.ac.th, \textsuperscript{2}Worasit.cha@dpu.ac.th

Corresponding Author: Pijitra Jomsri

https://doi.org/10.26782/jmcms.spl.9/2020.05.00006

Abstract

The tourist information recommendation system is useful for both tourist’s them-selves and tourist operators. This recommendation system can support tourists to spend less time searching for tourist attraction information and also be a channel for public relation to create incentives for tourists to use the services. User profiles is an important part of recommendation system that is responsible for finding the users’ interest and is a good representative for each tourist. However, creating a user profile to suitable each user in the tourist information recommendation system is still considered as challenging due to insufficient data collection. In addition, the use of social networks at present is becoming increasingly popular and is a source of information that has many users which can be extracted to represent the interests of each user. Therefore, this research has studied the recommendations for creating a user profile for the tourism information recommendation system in Thailand by using ATRU model to create User Profiling.

Keywords: User profiler; recommender system; travel recommender system;

I. Introduction

Nowadays, technology and communication has progressed and has been used extensively until becoming an important tool in conducting business both domestically and internationally. With an internet network as a tool that minimizes the size of the world, people can communicate and search information easily and quickly. This trend can be seen from all over the world with increased internet usage rates. Thailand has 16.1 million internet users or 24.4 percent of the entire population. Due to the increasing number of internet users, the industry and businesses are increasingly turning to the importance of electronic commerce [VI].
Online tourism business is a business that consumers search for information and use the internet service. Studying internet user behavior in 37 countries such as the United States, United Kingdom, and France found that consumers prefer to book accommodations, tours, and air tickets as the top 3 products and services that can make the highest income.

Tourism that has grown exponentially has increased the number of websites that support tourism whether it is a website that supports tourism-related decisions from static information or real time focusing on supporting the travel planning of users which considers transportation factors, as well as other factors for choosing attractions and services. In general, these systems will allow users to specify the basic characteristics of the user, the needs or interests of the tourism system, and then compared with the data in the Catalog electronics to present tourist-related attractions [XIII].

In search of tourist destination, users may sometimes use the travel website service that the government or organization has prepared which provides only basic information about tourist attractions. However information a fixed and recommended according to the nature of the place, without taking into account the unique characteristics of the user. At the same time, there are online social networking sites that provide travel opportunities which allow users to review and rate various attractions such as www.tripadvisor.com resulting in making a difference or ranking places in each group as an introduction from other people's experiences which make users more reliable. Moreover, in the ranking order of users vote, sometimes there may be problems in the matter of more than 1 place received equal votes; therefore, it needs to use other factors to support. However, Ricci and Werthner said that the introduction of attractions based on that individual's interests was like solving complex problems [XIV]. There were no exact methods or guidelines because the interest of the person might have varied according to each nation or experience. For example, Thailand may be an interesting tourist destination for Japanese people but not interesting for Egyptian. For this reason, it is a challenge to develop a system that introduces tourist attractions which truly meet the interests of users. Therefore, this research has studied the model for developing a user profile that is suitable for users who recommend tourist attractions in Thailand, taking into account both explicit and implicit factors of the users that affect the selection of tourist destinations to allow users to be guided to the most suitable and appropriate tourist attractions.

II. Related Work

A user profile is a representation of user’s interests and preferences that is used to verify to what extent news stories are relevant to a particular user. The profiles are built for each individual user, are regularly updated, and describe topics, newscategories and relevant features of the users. In principle, there are two types of user profiles, profiles based on implicit feedback and profiles based on explicit feedback [V]. Implicit user profiles are automatically extracted by the recommender systems themselves and may or may not be a correct representation of users’ interests. In general, implicit feedback methods assign relevance scores to user actions on news articles like sharing and bookmarking. Explicit profiles are approved by the users, but
tend to be less detailed than the implicit. The user selected categories in Flipboard form simple explicit user profiles. However, adopting numeric or symbolic scales increases the users’ cognitive load and may not be adequate for capturing emotions or attitudes towards the news. For mobile news recommendation it seems difficult to require that the user enter and regularly update extensive representations of her interests, though, and more advanced techniques for profile construction on the basis of implicit feedback are needed.

Some system creates user profile by combining explicit feedback and automatic learning [XV]. After building an initial interest category hierarchy on explicit feedback on a number of articles, the system analyzes user feedback from news sessions or update existing in the user interest. A similar approach is taken depend less on explicit user feedback [XV].

Some researcher developed an approach in DailyLearner for interpreting implicit user feedback [II]. A user click on the headline of an article is taken as a signal of interest. If the user is requesting more pages of the story, the score will be gradually increased when all pages have been consulted. Similarly, a skipped article is assumed to be of no interest and is given a negative score that is subtracted from the system’s prediction score for the article. All these scores are combined into a user profile that lists weighted informative words typical for the user’s interests and preferences.

The domain of tourism and traveling is very appealing for recommender systems research [III]. The commercial value of tourism is huge and e-Tourism is becoming increasingly popular. Moreover, people usually spend a considerable amount of time at planning their travel, and asking others for advice before making a decision. However, the particular characteristics of this domain provoke the appearance of novel problems and the need of developing new techniques to solve them [III]. Some researcher created personalized recommender system in the tourism field by use hybrid recommender system model [VIII]. Another researcher uses semantic expansion in combination with a standard user preference algorithm for content-based recommendation [IV]. The researcher perceivethat learned profiles tend to be dominated by the main characteristics of user’s preferences, preventing the recommendation engine from recommending news that are related albeit not directly addressed by the profile. Their solution is to use an ontology to include additional weighted concepts in the profile that are related to the original learned concepts or terms. Then include this expanded profile, which is a weighted set of concepts from the user’s latest interactions with the system to produce a contextualized version of the user’s preferences that filters out topics that are out of focus. Model for create user profile vectors which express users’ interests in specific news categories over time [X]. For each use they record the distribution of clicks and associate these click rates with categories on a monthly basis. This allows them to assess for every user the proportion of time spent on reading news from each category as well as to reflect on the development of her interests from one month to another.

The main of recommender systems in the part of tourism was already stated [XIII]. This research explained that content-based systems is suitable for recommending travel destinations. Travel destinations can be portrayed by rather stable concepts, and
as such a good knowledge base can be reused by the rent engines [XIV]. This approach allowed users to alter and tweak these existing travel plans and save them in the system. Activities such as traveling, dining, or enjoying the nightlife are usually social activities, which are carried out in groups of people (couples, families, colleagues, friends); thus, it is necessary to take into account the preferences of all the participants when providing recommendations [XI]. This introduced the need for intelligent systems generating group recommendations covering the preferences of all group members.

III. Proposed Approach

Tourism Recommender system for personalization is a system that offers tourist information in various ways such as tourist information travel plan travel package information to tourists according to different interests in each person. The system has the main structure as follows: 1) End users 2) Data collection process 3) Modeling the user profiler which is to find the conclusion, interest and keep up with the interest that changes to the user through multiple sources 4) Tourism Attractions Recommendation is a guide to attractions by using interest information based on the user profiler model.

End users consist of tourists, administrators, target tourists and experts. Tourists refer to those who have experience in tourism and can provide brief history and opinions related to tourist attractions. The system administrator is responsible for creating a ranking model for tourist attractions. The generated model can be changed if new information is imported from tourists. The next group of users is tourists who want to recommend tourist attractions which is called "Target tourists". And the last group is travel experts acting in scoring tourist attractions to be the score used in the ranking together with collected tourist information.

The data collection process is the procedure for collecting information of user in tourist system and tourist location which were collected from experts to be used in the development process of ranking tourist attractions.
Fig. 1: Framework for Tourism Recommender system

The system for personalized recommender system requires summary information about the user's interest to be used in introducing travel information that meets the needs of users which consists of 3 components 1) Type of user profile 2) Information Acquisition, and 3) User Model Adaptation which will analyze how to use the method for user interest information which show in Figure 2. In addition users who start using the system to select a category and create an initial user model, such information will be sent to the User Model Adaptation for update for users who have already used the system. The system will update according to the information that users are interested in to better meet the needs of users.

Fig. 2: Component of User Profile Model

It exists two types of data collected and stored in the user profile, explicit and implicit Data. Explicit data are Information that the user explicitly entered for example, the user name, gender, or by explicit feedback such as ratings. The user profile was created in order to facilitate the extraction of the user personal information. The user is requested to register and fill information in few forms with personal information such as login, age, gender, origin, region, travel style. Their area of interest in tourism such as historical places, museum etc. The user is also requested to provide trip date.
Based on the information about accompanying persons with the user, their area of interest perfect destination to the user can be identified.

In addition, implicit data is determined by analyzing user behavior and the actions user performs during his interaction, the implicit feedback are examples of implicit data that must be incorporated in the user profile.

**Table 1:** User Interest Level

| Input                                                                 | Output               |
|----------------------------------------------------------------------|----------------------|
| All check-ins data of a user                                         | Interest level $I_u(j)$ in each category. |
| Take off duplicated check-ins from the same day                      | 1: Take off duplicated check-ins from the same day. |
| Take off any check-ins which are not attractions                      | 2: Take off any check-ins which are not attractions. |
| Classify check-ins                                                     | 3: Classify check-ins. |
| If all classified check-ins of a user $\geq$ threshold                | 4: If all classified check-ins of a user $\geq$ threshold. |
| Calculate level of interest $I_u(j)$                                 | 5: Calculate level of interest $I_u(j)$. |
| Else                                                                  | 6: Else. |
| Define level of interest $I_u(j)$ with average value from popularity | 7: Define level of interest $I_u(j)$ with average value from popularity. |
| Return $I_u(j)$                                                       | 8: Return $I_u(j)$. |

There is scenario for extracting social network check-in data to detect user preferences: adequate and inadequate information. Adequate information refers to users with check-in data greater than the threshold, which is chosen experimentally to find user interests while inadequate information is the number of check-ins of a user that are less than the threshold. To analyze user interest based on personal data, the computation scheme of user interests can be defined as (1), where $I_u(j)$ is an interest level in a category $j$ of the user $u$, $n_j$ is the number of check-ins for a category $j$ and $I_u(j)$ is normalized in the range of 0 and 1.

$$I_u(j) = \frac{n_j}{\sum_{i=1}^{5} n_i}, \quad 1 \leq j \leq 5$$

Therefore, this research a heuristic user profile by using activity, travel style, review score and user Interest or ATRU model.

### III. Experimental Setting and Result

In this section the implementation of user profile for tourism recommendation system and experiment were described.

**Data Set and Pre-Processing**

Data set were crawled from TripAdvisor (www.tripadvisor.com) which is social network for travel and review through a web crawler. The engine is able to navigate through web pages and convert information presented in term of HTML form into a structured data. An example of data set was show in Figure 3. This web crawler is able to obtain product’s information, user’s information and user’s review and evaluation of system. Data is included Users information (id, login, age, gender, origin, region, travel style). Activities (auto id, activity name, activity category, activity price, location, POI). Rating (activity id, user id, rating, review).
Fig. 3: Example of data set from TripAdvisor
To perform the experiment data were collected from 30 users with different age and background. The subject tests were given a list of 90 locations from Bangkok, which are used by the TripAdvisor application. Subjects were asked to rate given location for interest which are familiar with from 1 to 5 stars. The data was used to evaluate each module separately and determine the final weights. The goal of each module is to accurately predict the rate of user interest place as would be given by the current user. Mean Absolute error (MAE) calculated result by as defined by equation 2 [VI].

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |x_i - y_i|
\]

where \( x_i \) is the predicted rating value by the \( i^{th} \) tourist, \( y_i \) is the real rate and \( n \) represents the number of activities with real rate in testing set. The lower MAE is, the more accurate is the recommendation result. This paper has used also a normalized version of the MAE to express errors as percentages of full scale [VII], this metric is defined by:

\[
NMAE = \frac{1}{n} \frac{\sum_{i=1}^{n} |x_i - y_i|}{(r_{max} - r_{min})}
\]

where \( r_{min} \) is the ratings minimum value and \( r_{max} \) is the rating maximum value. The Root mean squared error (RMSE) metric was applied [I], which is widely used in the literature for evaluating recommender systems which defined by:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2}
\]

The RMSE emphasizes large errors by shedding a more severe penalization over them compared to the other metrics.
Table 2: Example of Ratings

| User No. | Activity No. | Rate Score |
|----------|--------------|------------|
| 1        | 10           | 2          |
| 3        | 8            | 4          |
| 3        | 8            | 4          |
| 7        | 1            | 5          |
| 7        | 1            | 5          |
| 15       | 2            | 5          |

IV. Result and Discussion

In this section, present the experimental results of research method in generating ratings. In order to obtain the best similarity measure to be applied in the CF approach, the 5 subsets were performed experiments. The researcher calculates for each data by using MAE measurement and also the MAE criterion for the averages was computed. Table 3 represents the obtained results.

This research note through Table 3 that the Euclidean distance have the lowest value of MAE among the five testing subsets compared to the Cosine. Also the result can note that overall, the Euclidean distance Coefficient CF achieves mostly better results according to our data set comparing to the Cosine approaches with an average of 0.81 for the Euclidean distance 0.886 for the Cosine. Given these results, the Euclidean distance as a similarity measure of our CF approach.

Table 3: MAE of CF with two different similarity measure

| Subsets   | Euclidean distance | Cosine  |
|-----------|--------------------|---------|
| Subset 1  | 0.79               | 0.89    |
| Subset 2  | 0.89               | 0.92    |
| Subset 3  | 0.8                | 0.93    |
| Subset 4  | 0.78               | 0.84    |
| Subset 5  | 0.79               | 0.85    |
| Average   | 0.81               | 0.886   |

Table 4: The result of the assessment

| Question List       | Overall Average | Reliability |
|---------------------|-----------------|-------------|
| Efficiency of use   | 4.85            | 0.35        |
| Effectiveness       | 4.53            | 0.59        |
| Flexibility         | 3.95            | 0.47        |
| The ability to learn| 4.32            | 0.55        |
| The satisfaction of users | 4.26       | 0.24        |
| Average             | 4.382           | 0.44        |

The rating of the speed rating and evaluating the usability of the system is assessed by the level of satisfaction with the user system, as well as allowing users to comment on
additional system usage. The satisfaction level score is classified as the highest, most, moderate, least and least, then consider the average level of satisfaction towards the usability of the 5 different systems.

V. Conclusion

This research has developed the creation of a User profiler for tourism recommendation systems in Thailand with the objective in order to design and develop models that are used to introduce attractions based on personal interests which is the prototype of the User Profiler that has been developed and applied to the tourism recommendation system in Thailand. In the assessment of that system, this research evaluates the accuracy which considers MAE with two different similarity measure that the tourists arranged with the orders of the places that the models suggest. The results of the assessment showed that users are overall satisfied with the system's capabilities at a higher level in all aspects which is true according to the hypothesis. Moreover, from MAE testing the result of this study shows that ATRU model can recommend attractions and present to the users appropriately. In addition, this system can provide users with overall satisfaction at the highest level and can use it correctly, meet the goal, and can achieve.

VI. Acknowledgment

The authors would like to thank Thailand Science Research and Innovation (TSRI) for scholarship support.

References

I. Bennett, J., C. Elkan, B. Liu, P. Smyth, and D. Tikk, “Kdd cup and workshop 2007”, ACM SIGKDD Explorations Newsletter, vol. 9, no. 2, pp. 51–52, 2007

II. Billsus, D. and Pazzani, M. J., “User Modeling for Adaptive News Access. User Modeling and UserAdapted Interaction,” 10, pp. 147-180, 2000.

III. Buhalis, D., Law, R., “Progress in information technology and tourism management: 20 years on and 10 years after the internette state of etourism research,” Tourism Management 29(4), 609 – 623 (2008). DOI http://dx.doi.org/10.1016/j.tourman.2008.01.005. URL http://www.sciencedirect.com/science/article/pii/S0261517708000162

IV. Cantador, I. Bellogin, A. and Castells, P., “Ontology-Based Personalized and Context-Aware Recommendations of News Items,” In Proceedings of the 7th International Conference on Web Intelligence, pp. 562-565. IEEE. 2008
V. Gauch, S., Speretta, M., Chandramouli, A., and Micarelli, A., “User profiles for personalized information access,” The adaptive web, pp. 54-89. Springer, 2007.

VI. J. S. Breese, D. Heckerman, and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering”, The Fourteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., pp. 43–52, 1998

VII. K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, “Eigentaste: A constant time collaborative filtering algorithm,” information retrieval, vol. 4, no. 2, pp. 133–151, 2001

VIII. Kbaier M. E. B. H., Masri, H.; KrichenS.; “A Personalized Hybrid Tourism Recommender System”, 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), 2017

IX. Kim, H. R. and Chan, P. K., “Learning implicit user interest hierarchy for context in personalization,” In Proceedings of the 8th international conference on Intelligent user interfaces, pp. 101–108. ACM, 2003.

X. Liu, J. Dolan, P. and Pedersen, E. R., “Personalized news recommendation based on click behavior,” In Proceedings of the 15th international conference on intelligent user interfaces, pp. 31–40. ACM, 2010.

XI. Moreno, A., Sebastia’, L., Vansteenwegen, P., “Tours’15: Workshop on tourism recommender systems,” the 9th ACM Conference on Recommender Systems, RecSys’15, pp. 355–356. ACM, New York, NY, USA, 2015

XII. Moreno, A., Sebastia’, L., Vansteenwegen, P.: Tours’15: Workshop on tourism recommender systems. In: Proceedings of the 9th ACM Conference on Recommender Systems, RecSys’15, pp. 355–356. ACM, New York, NY, USA (2015). DOI 10.1145/2792838.2798713. URL http://doi.acm.org/10.1145/2792838.2798713.

XIII. Ricci, F., “Travel recommender systems, “ IEEE Intelligent Systems”, 17(6), 55–57, 2002

XIV. Ricci, F., Werthner, H., “Case base querying for travel planning recommendation,” Information Technology & Tourism 4(3-4), 215–226, 2001

XV. Singh, S., Shepherd, M., Duffy, J. and Watters, C., “An Adaptive User Profile for Filtering News Based on a User Interest Hierarchy,” In Proceedings of the American Society for Information Science and Technology, Volume 43, Issue 1, pp. 1-21, 2006.