DETAILED STRUCTURE APPLICABLE TO HYBRID RECOMMENDATION TECHNIQUE FOR AUTOMATED PERSONALIZED POI SELECTION

Phuengjai Phichaya-anutarat* and Surasak Mungsing

Information Science Institute of Sripatum University (ISIS)
Sripatum University, Chatuchak, Bangkok 10900, Thailand

ABSTRACT

The main purpose of this paper is to present the detailed descriptive structural processes and algorithms that are applicable to a hybrid recommendation technique (the combinations of content-based, collaborative filtering and demographic techniques) to address personalized Point of Interest (POI) selection problems in the field of tourism. The technique emphasized on the recommendation of the POI, based upon the highest personalized tourist’s preferences and interests. With this technique, POIs could be automatically selected based on only preliminary tourist’s information through the well-designed queries. There were three parts of concerns in the presentation. The first part is the design of information query, in which the tourist would be asked, in particularly, for his/her preferences or interests and trip constraints. The second part is the utilization of tourist’s information obtained from the previous part to obtain POI scores, using each recommendation technique. The outcome was expressed in terms of the score that assigned according to the POI significant. The third part involved clustering of the possible POIs. The final result yields an appropriate trip area that most satisfied tourist’s highest personalized interests and requirements. The proposed procedures and algorithms can be implemented for automated personalized POI selection.

Keywords: Clustering, Hybrid recommendation technique, Personalized POI selection, Tourism
1. INTRODUCTION

Dating back to the past two decades approximately, there are many researches applied on the recommender technologies found in scientific literature. Recommender systems have become an important research area since the first related research paper on this field was appeared in the mid-1990s. Significantly, it can be observed that they have increased rapidly in recent years and still developed continually in various fields of application. This is confirmed by and stated in the work of Park et al. [11], who comprehensively reviewed the related papers with 210 articles published between 2001 and 2010 on recommender systems research that can be categorized into eight application fields and together with eight data mining techniques.

Because one of the key features of the recommender systems is the ability to consider the needs and preferences of the user who has an individual preference and interest in order to generate personalized recommendations or suggestions, with the advantages of recommender system it can primarily help users to access complex information spaces in response to user needs and preferences. Therefore, they are designed to address many of the problems by offering users a more intelligent approach to navigating and searching complex information spaces or items of interest.

Based on the recommendation techniques that have been used widely in general recommender systems, they come in three basic flavors [13]. The first one is the Content-Based (CB). The second one is the Collaborative Filtering (CF) and the last one is the demographic technique. However, some research works have been combined these techniques or at least two of them in their applications, and known to be the hybrid technique. It is able to be noted here that the term ‘hybrid recommendation technique’ or shortly, ‘hybrid technique’ has become synonymous with the problems of using at least the combination of two techniques.

Nowadays, there is a vast amount of recommendation techniques that employed in several recommender systems. However, these techniques are still quite vulnerable to some limitations and shortcomings related to recommender environment in each technique. Considering organization of document for knowledge management, Liu et al. [9] presented novel document recommendation methods, including content-based filtering and categorization, collaborative filtering and categorization, and hybrid methods, which integrate text categorization techniques, to recommend documents to target worker’s personalized categories. From the experiments that have conducted to evaluate the performance of various methods using data collected from a research institute laboratory, it can be conclude that the proposed methods can proactively provide knowledge workers with needed textual documents from other workers folders in which the hybrid methods outperform the pure content-based and the collaborative filtering and categorization methods.

de Campos et al. [4] combined two traditional recommendation techniques which are content-based and collaborative filtering in order to improve the quality of the recommendation because both mentioned techniques have their certain disadvantages. Therefore, they have proposed a hybrid recommender model based on Bayesian networks which uses probabilistic reasoning to compute the probability distribution over the expected rating. From the test results, it is remarkable that the proposed model is versatile and can be applied to solve different recommendation tasks such as finding good items or predicting ratings. Another work that has been combined content-based and item-based collaborative
filtering to eliminate the most serious limitations of collaborative filtering alone was made by Barragáns-Martínez et al. [1], who described the design, development, and startup of a TV program recommendation system. The proposed hybrid approach provided all typical advantages of any social network, such as supporting communication among users as well as allowing users to add and tag contents, rate and comment the items, etc. Importantly, a number of experiments were set up and tested with real users in which very much positive feedback have received.

Due to information overload, the importance of personalized recommendation systems for online products and services is rapidly growing. Such systems allow buyers to find what they want without wasting their time and also enable sellers to provide buyers with items they are likely to purchase, thereby furnishing benefits to both parties. Choi et al. [3] implemented a hybrid online-product recommendation system, which integrates Collaborative Filtering-based recommendation (CF) using implicit ratings and Sequential Pattern Analysis-based recommendations (SPA) for improving recommendation quality. Referring the obtained results of four experiments, they have contended that implicit rating can successfully replace explicit rating in CF and that the hybrid approach of CF and SPA is better than the individual ones. In social annotation systems that enable the organization of online resources with user-defined keywords, collectively these annotations provide a rich information space in which users can discover resources, organize and share their finds, and collect to other users with similar interests. However, these systems led to information overload and reduced utility for users. Therefore, Gemmell et al. [6] introduced a linear-weighted hybrid recommendation algorithm and showed how this technique serves to combine multiple complementary components into a single integrated model that provides the most flexibility considering the unique characteristics across different social annotation systems. In additions, they also discovered the performance of the algorithm on six large real-world datasets, and on two different variants of the resource recommendation task.

Karden and Ebrahimi [7] presented a novel approach in hybrid recommendation systems based on association rules mining technique for content recommendation, which can identify the user similarity neighborhood from implicit information being collected in asynchronous discussion groups. The experiments carried out on the discussion group datasets have proved a noticeable improvement on the accuracy of useful posts recommended to the users in comparison to content-based and the collaborative filtering techniques as well. Recently, Moscato et al. [10] dealt with a new vision of multimedia recommender systems based on a novel paradigm that combined both analysis of user behavior and semantic descriptors of multimedia objects in digital ecosystems (system for e-business applications) in which past behavior of users in terms of usage patterns and user interest can be expressed in an ontological shape. Moreover, the use of a knowledge based framework can reduce the complexity of generating recommendations with the aid of semantics.

The rest of this paper is organized as follows. Section 2 describes the statement of problem considered here and discusses previous research works that related in the field of tourism. Section 3 presents the design of framework and provides essential details to tourist information. Reviewing recommendation techniques that used in this paper is also given together with the description of POI clustering concept. In Section 4, the
details of computing POI score are explained step-by-step through an illustrative example. Finally, Section 5 presents our conclusions and outlines future lines of possible research.

2. PROBLEM STATEMENT AND RELATIVE WORKS

As mentioned above, available recommendation methods were applied in many problem fields with applications such as automatically matching user’s likes to TV programs and recommending the ones with higher user preference [1], personalized recommendation systems for online products and services in many online shopping malls [3], recommendation problem for organization of online resources with user-defined keywords in social annotation systems [6], document recommendation for knowledge management system [9], and recommendation system for e-business applications in digital ecosystems [10]. It is obvious that these problems are considered and developed because of a huge amount of complex and heterogeneous data that produced and transferred on the Internet. Furthermore, several hybrid approaches have been proposed and successfully used in their vast majority concerning combinations of content-based and collaborative filtering techniques since there are still some other limitations of either content-based or collaborative filtering techniques. In order to produce high quality recommendations and its performance degrades with the increasing number of users and programs, new recommender system technologies are needed to quickly produce high quality recommendations, even for very large-scale problems.

A closely related works with respect to tourism problems was explored and reviewed as follows. Souffriau et al.[16] presented a new approach that enables fast decision support for tourists on small footprint mobile devices, to calculate tourist routes in a dynamic context by combining techniques from the field of information retrieval with using techniques from operational research. Documents related to physical locations are indexed by using the vector space model where the resulting locations and their personalized interest scores form the basis for formulating a tourist trip design problem, in which the total score has to be maximized without exceeding a given distance or time budget.

Schiaffino and Amandi [14] described and presented an expert software agent in the tourism domain, which combines collaborative filtering, content-based user profiles and demographic information to recommend tours and package holidays to users. The combination of three approaches called hybrid approach takes advantage of the positive aspects of each technique and overcomes the difficulties by each of them when used in isolation. Noted from the results obtained, the precision of the recommendations is higher for the hybrid technique than with each method used separately. Another expert system that involved in tourism problem was made by Vansteenwegen et al.[21] who introduced a tourist expert system that allows planning routes for five cities in Belgium and implemented as a web application. This system takes into account the interests and trip constraints of user through a small questionnaire and matches these to a database of locations in order to predict a personal interest score for each POI that tourist can visit within the constraints.
In case of Multi-Period Orienteering Problem with multiple Time Windows (MuPOPTW) which is a problem encountered by sales representatives when they want to schedule their working week, was studied and developed by Tricoire et al.[19]. Because this problem is too hard for contemporary commercial solvers, at least for interesting size instance, they therefore proposed an exact algorithm for a route feasibility subproblem, featuring multiple time windows and also embedded it in a Variable Neighborhood Search method (VNS) to solve the whole routing problem and provide good solutions in reasonable amounts of time. However, a variant application of the orienteering problem has previously collected and reviewed by Vansteenwegen et al.[20].

Consider the problem of group recommender system, Garcia et al.[5] presented a recommender system that applied for tourism based on the tastes of the users, their demographic classification and the places they have visited in former trips. The proposed system is able to offer recommendations for a single user or a group of users. For the single user recommendations, they can be computed according to the user preferences with the use of hybrid recommendation technique that combined the demographic, content-based and general likes filtering techniques. In case of a group of users, group preferences can be obtained from the individual preferences through the application of the intersection and aggregation mechanism.

Recently, Batet et al.[2] applied a recommender system in e-Tourism for helping tourists to select appropriate destinations. The system is designed within the following three main goals. The first is to provide an easy and ubiquitous access to the desired information about tourist attractions. The second is to provide proactive recommendations of attractions using a hybrid recommendation system. Finally, the third is to implement a high degree of dynamicity and flexibility in which the system can adapt to changes in the activities and incorporate new information at execution time. Significantly, the system can be taken into account the user profile and preferences, the location of the tourist and the activities, and also the opinions of previous tourists. Shinde and Kulkarni [15] proposed and described a personalized recommender system using Centering-Bunching Based Clustering (CBBC) algorithm in which the system is consisted of two main phases. In the first phase, opinions from the users are collected in the form of user-item rating matrix, and further clustered offline using CBBC into predetermined number clusters and stored in a database. In the second phase, the recommendations can be generated online for active user using similarity measures by choosing the clusters with good quality rating, which leads to get further effectiveness and quality of recommendations for the active users.

The details that given in the following stages with respect to this section are the sequel of the previous work that presented by Phichaya-anutarat and Mungsing [12] which has proposed the concept of hybrid recommendation technique to assist the tourist in order to select appropriate the Point of Interest (POI) according to the personalized tourist’s preference and interest. Nevertheless, the objectives are therefore concentrated to explain and demonstrate the evaluation of personal interest POI locations represented in terms of the assigned score.
3. FRAMEWORK DESIGN AND RECOMMENDATION TECHNIQUES

In this section, the essential components that involved and used in details of evaluating the POI scores for each recommendation technique are given hereafter. Section 3.1 describes the structure of the personal tourist information. This tourist information will...
further be considered and matched to the pre-determined POI recommendations in terms of significant score. Before proceedings further to employ the recommendation techniques in order to suggest the POIs, Sections 3.2 to 3.4 briefly explain three well-known recommendation techniques which are the content-base, collaborative filtering, and demographic techniques [12,13], to be used in order to predict personal interest scores for each POI location in accordance with each recommendation technique. Section 3.5 introduces the clustering technique in order to find the groups of POIs that concerned within trip constraints of tourist as specified in Section 3.1. These can be overall represented in Fig.1 and explained in the following subsections below.

![Figure 2 POI Information](image)

**3.1 Tourist Information**

The tourist which means to active user in the subsequent sections will be asked for and has to enter a level or degree of his/her personal preferences or interests for POI’s type and category. Each of tourist’s POI type and category membership, totally not, a bit, mostly, and absolutely, is quantified as 0, 1, 2 and 3, respectively. However, it should be kept in mind that each of POI that belongs to exactly one GPS coordinates and one type (e.g., temple, museum, churches, market, etc.). Nevertheless, a POI can be identified more than one category (e.g., archaeology, religious, nature, local place, floating market, etc.) or otherwise, this means that there can be more than one category classified for a single POI [12, 21]. Particularly, the trip constraints that composed of date and duration, travel time, recommendation ratio, and radius of trip area are also needed to fulfill completely. Based on these constraints, the set of POIs that can be visited within trip constraints are met. Furthermore, all this tourist information should not take much more than a minute.
3.2 Content-Based Recommendation Technique

This technique is based on the intuition that each user or tourist exhibits a particular behavior (preferences and interests) under a given set of circumstances in which this behavior will be repeated until satisfy the condition of similar situations. Importantly, it can also learn a model of the user interests based upon the features present in items the user rated as interesting either by implicit or explicit feedback. Thus, to perform this kind of recommendation, a user profile contains those features that characterize a user interests, enabling agents to categorize items for recommendation based on the features they exhibit. For simplicity, it is then required to build a user profile that stores the degree of interest (i.e., a score) on each of the different criteria. However, such information can also be extracted by fill-in forms but, as the set of characteristics can be quite large.
3.3 Collaborative Filtering Recommendation Technique

One of the well-known recommendation techniques that have been used widely in the past by many researchers is the collaborative filtering. It is based upon the idea that people within a particular group tend to behave alike under similar circumstances. The behavior of a user can then be predicted from the behavior of other like-minded people. Therefore, this technique requires several ratings from the users before the system may begin to make recommendations. As a result, a user profile comprises a vector of item ratings, with the rating being binary or real-valued. It can further be noted that it is to predict the score for an item which has not been rated by the active user in order to recommend this item. Comparison between the ratings of the active user and those of other users using some similarity measure, the system determines users who are most similar to the active one, and makes predictions or recommendations based on items that similar users have previously rated highly.

Figure 4 Destination-Map According to Trip Constraints
3.4 Demographic Technique

Following the application of demographic technique, it assigns each user to a demographic class based on their user profile so that it is mainly aimed at categorizing users based on their personal attributes as belonging to stereotypical classes. This will be used further to form justifications for recommendation since users have received recommendations according to the group in which they were classified. Therefore, a user profile is a list of demographic features representing a class of users in which representation of demographic information in a user profile can vary greatly. It is, however, notable that this demographic technique makes nearest-neighbor or other classification and clustering tools.
3.5 POI Clustering

After obtaining the POI scores for each possible POI selection from the above recommendation techniques, the use of advantages that found in clustering concept [8,13,15] can be extended to perform those mentioned POIs efficiently for finding appropriate groups of POI according to trip constraints as previously defined by user in Section 3.1. For this purpose, the details of clustering application will be explained in Section 4.3.

4. TECHNICAL APPROACH

In view of Fig.1 which presented an overview of structural process of the proposed approach, it is seen that the main process of this approach can be separated and categorized into three parts. The first is the related data and management. The second is given for the input data. Finally, the third is an automated personalized POI selection and trip area. However, all these three parts are also inherently consisted of sub-procedure which can be described as in the following subsections below.

4.1 Related Data and Management

In the first part of the structure proposed here that is concerned to the related data and management. There are two components which are the database manager and database. They can be explained in details as follows.
Figure 7 POI Score Based on CB-Recommendation Technique
4.1.1 Database Manager

Database manager is the middleware that connected between the databases and other parts that required data relatively. Thus, the major function of database manager is to receive and send the data. In additions, it has also to search, coordinate and send the data to other parts that have involved and requested between POI information manager and tourist information manager.
Within the data manager, it is consisted of POI information manager and tourist information manager. In part of POI information manager, its function is to manage POI information containing POI details and recommended POI which already collected in the POI database. This POI information will then be input and updated by the system administrator only for which to be used as a supporting data to other parts such as automated personalized POI selection and trip area (APPOISTA). This is to propose to be as
a part of tourist feedback database. Another part of data manager is the tourist information manager. This part is assigned to manage the tourist information that consists of either tourist profile or tourist feedback. The tourist profile will be kept into tourist profile database and tourist feedback is for tourist feedback database. It is, however, noticing that the tourist information can also be used as a supporting data for other parts such as APPOISTA in the process of hybrid recommendation technique for selecting personalized POI.

**Figure 10** Summary of POI Score Based on each Recommendation Technique

Tourist profile (e.g. tourist interest and preference and personal data) is the input data that contained in the tourist profile database. It can be input directly from the tourist for the first time only, but for the next time it will be updated according to the present data of tourist. The later can be performed either direct or indirect methods. In case of the direct
method, the data that have contained in the tourist profile database will be requested through the tourist information manager, and are updated by tourist. For the indirect method, the data obtained from the tourist feedback and collected in tourist feedback database are requested to update. It can be remarked that the tourist feedback are the input data which have input and updated by individual tourist. Therefore, tourist should give always his/her feedback involving appropriate trip area with selected personalized POIs. This purpose is to make the data up-to-date for his/her information.

![Summary of POI Score Based on Hybrid Recommendation Technique](image)

**Figure 11** Summary of POI Score Based on Hybrid Recommendation Technique

### 4.1.2 Database

The second component in the related data and management part is the database. This component has three main essential databases, namely, point of interest, tourist profile, and tourist feedback. The detailed description of them is then given in the followings below.
**Point of Interest (POI) database** is designed to keep and to collect the POI information as shown in Fig. 2. There are two main data information in this mentioned database. The first is POI details and the second is recommended POI.

POI details can be defined to be specific information of each POI that provided the essential information or details for identifying POI. Their details can be classified as POI name, GPS coordinates, type, and category. However, it is notable that each of POI can only be identified into one type, but unnecessarily for category. Additionally, a calendar of opening and closing days, a calendar of opening and closing hours, and average visiting duration are also included. All these will be used to consider the POIs selection according to tourist’s trip constraints.

Recommended POI, sometimes known to be POI as “not to be missed”, is the favorite POI that should be proposed, and visited by tourist. However, a number of recommended POI is not limited depending on the recommendation of system administrator.

**Tourist profile database** has a function to collect the specific data or information of each tourist which consists of the tourist’s personal data, interest and preference, and travel history that illustrated in Fig. 3. All these are input and updated by tourist as described previously in Section 4.1.1. In part of personal data or demographic profile, it is an individual data of tourist that represented in form of demographic. This typically involves gender, age, education degree, and income. Moreover, the present work is also interested in address and job. This demographic profile is important in the process of automated personalized POI selection using demographic recommendation technique. Next, consider the interest and preference which can be defined to be the degree or level of tourist’s interest and preference to type and category of POI. They are quantified by the scores of 0, 1, 2, 3 according to totally not, a bit, mostly, and absolutely, respectively. Both type and category of POI have to be specified with a single score only.

The last database is the **tourist feedback database**. The history of POI including its score of interest and preference for each tourist are contained in this database in which each POI has been selected in appropriate trip area with selected personalized POIs in the previous time.

### 4.2 Input Data

**4.2.1 POI Information**

To consider this information, only the system administrator is supposed to input and update the POI information to be kept up-to-date by sending information through POI information manager into POI database as demonstrated in Fig.2. In case of a new POI that has not yet contained in POI database, system administrator will have to input that POI details and recommended POI for the first time. In order to describe the POI to tourist clearly, with using the GPS coordinates of each POI, it can make a possible map that is contained inherent details of POIs. This is illustrated in Fig.4.
Data concerning trip constraints, interest and preference, and personal data as seen in Fig. 3 are considered here. The trip constraints are the conditions or limitations of travel. They are set by the tourist, which have different details in each travel so that they are not kept in the database. Trip constraints can be divided into four parts as follows: date and duration, travel time for each day, recommendation ratio (%), and radius for trip area. The details of date and duration and also travel time will be transformed to a condition to be used for removal of POIs that cannot be satisfied as seen in Figs. 5 and 6. This is reducible computational time in the process of APPOISTA.

In part of recommendation ratio (%), it is introduced to further use in APPOISTA process which has a purpose for assigning the confidence of recommendation in terms of some factors. The factors $k_{CB}$, $k_{CF}$, and $k_{D}$ represent the confidence of recommendations according to the content-based, collaborative filtering, and demographic techniques,
respectively. The value of radius distance shown in the radius of trip area part is the condition that defined to be the maximum distance cover in the trip area.

4.3 Automated Personalized POI Selection and Trip Area

The third part that has shown in Fig.1 for the structure overview will be considered in this section. It contains three main components in which the details of the first component have been described in the previous sections. Therefore, the remaining two components that are hybrid recommendation technique and clustering technique will be explained here. However, it is more convenient way to present them through the illustrations with their detailed explanations.

In the hybrid recommendation technique, there are three well-known techniques to be applied in the present work, namely, the content-based, collaborative filtering, and demographic techniques. All of them have demonstrated in Figs.7, 8, and 9, respectively. Consider Fig.7 that gives the results of POI with type and category scores based on the content-based technique. It is obvious that the notation “Y” is represented as the recommended POI by system administrator, otherwise “N” is not for. Also noting that the scores of each POI can be computed from the scores’ type and category as given previously in Fig.3. If a POI has a type score of 3, its type score will be multiplied with 3. In case of category score of 3, its category score will also be multiplied with 2. Otherwise, both for POI’s type and category are multiplied with 1. Since a POI is assigned to be the recommended POI, the additional bonus of 6 is awarded to that POI. Therefore, the total score for each POI can be determined by multiplying between type score and summation of category scores as demonstrated in Fig.7. By the same procedure, this descriptive score computation is also applied to both of collaborative filtering and demographic techniques that already illustrated in Figs.8 and 9, respectively.

Finally, each score that obtained from each technique will be adjusted by the recommendation ratio in terms of corresponding factor. The results are given in Fig.10. Thus, the total POI score can be determined by combination of those three adjusted scores according to each recommendation technique that completely shown in Fig.11.

The last component is involved with clustering technique [13]. Based on this technique, the appropriate trip area can be determined from the trip destination containing possible POIs that found in hybrid technique. In the present work, the clustering technique will be applied twice times for personalized POIs according to preference and interest and also trip constraints. The first POI clustering is based on all of recommended POI clustering concept in which the numbers (integer) of cluster are computed from the ratio of destination area and the desired trip area. This POI clustering is shown in Fig.12. With the use of previous results that are the predetermined clustering, the centers of each cluster will be specified to be the same centers in the second time of clustering. The size of all clusters is based on the radius distance given by tourist (see Fig.3). The score of each cluster is the summation of POIs score within trip constraints. Therefore, the highest score is then selected to be the finalized appropriate trip area as shown in Fig.13.
5. CONCLUDING REMARKS AND FURTHER RESEARCH

The principal aim of this paper has two-fold. The first is to describe the design and development of hybrid recommendation technique that applied in tourism problem for automated personalized POIs selection [12]. The proposed technique is a mixing of three well-known techniques which are the content-based, collaborative filtering, and demographic techniques. Based on these mentioned techniques in conjunction with the designed tourist information, it can firstly recommend the tourist to receive POIs that represented in terms of the assigned POI score. Further utilizing the advantages of
clustering technique, the appropriate trip area containing the possible POIs and satisfying the highest personalized tourist’s preferences or interests and also trip constraints can finally be determined effectively. To practical implement and clearly understand the computational process and algorithms, the second is to present the application of the proposed hybrid technique through an illustrative example which has explained the state of the art of this proposed technique. In summary, the authors believe that, to the best of authors’ knowledge, no previous proposal addresses the proposed technique and efficiently solves the introduced problem fields using the present technique.

In terms of future research, there are some points that can still be researched:

- Design of new methods for estimating POI score with weights or factors to adjust the importance of recommendations in each technique. The work presented by Schiaffino and Amandi [14] and the concept given in Souffriau et al.[16] can be extended to this purpose.
- Design of group recommendation technique for automated POI selection based on the ideas of group preferences such as aggregation and intersection as given in Garcia et al.[5]. With the authors’ opinion, it may use the clustering techniques [8,15] to offer recommendations for a group of users with slight modifications.
- Further applying the POI selection for personalized tourist trip planning so that it can highly refer to Vansteenwegen et al.[21]. Additionally, either iterated [22] or guided [23] local searches based upon various heuristic approaches can be adopted. Furthermore, other techniques which have been proposed by Taplin and McGinley [18] and Tang and Miller-Hooks [17] may also be extended to finding trip planning for personalized tourist.

REFERENCES


