International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Survey of Travel Package Recommendation System

Parnika Patil[1], V. L. Kolhe[2]

[1]Savitribai Phule Pune University, D Y Patil College of Engineering, Akurdi, Pune, India (411044)

Abstract: ***Recommender systems are like information filtering system that seeks to predict the 'rating' or 'preference' that user would give to an item. Recommendation systems have become extremely common in recent years and are applied in a variety of applications like music, movies, news, research articles, books, social tags, search queries and products in general. This paper focuses on their application in tourism. It focuses on recommendation techniques and a survey of different AI techniques for travel package recommendation system.***

Keywords: Recommendation System, Travel Package, AI techniques, Clustering

1. Introduction

The Recommender Systems are software tools and techniques providing suggestions for items for a user. The suggestions provided are aimed at supporting their users in various decision making processes, such as where to plan a tour for which season, what items to buy, what music to listen, or what news to read. Recommender systems have proven to be valuable means for online users to cope with the information overload and have become one of the most powerful and popular tools in electronic commerce field. Various techniques for recommendation generation have been proposed and during the last decade like content-based, user-based and item based collaborative filtering and hybrid recommendation system. Many of them have also been successfully deployed in commercial environments. Added to it there are many evolutionary methods that could be incorporated to achieve better results in terms of accuracy in prediction, handling various challenges of recommendation system like data sparsity, cold start problem, scalability and accuracy issues.

Recommendation system is very useful for both customers and providers. For the customers it will help to narrow down the set of choices, explore the set of options, find the things that are more interesting to user and discovers the new things [2]. In case of provider, it will help to increase trust and customer loyalty, increase customer conversation, sales, click through rates. It will provide good opportunity for promotion and obtain more knowledge about the customer.

The reminder of this review is structured as follows. The section II comments the recommendation methods employed by tourism recommenders, focusing on content-based, collaborative and hybrid approaches. The section III exposes the use of different AI techniques from different fields like optimization techniques and clustering approaches etc. and section IV summaries the paper.

2. Approaches to Recommendation System

Recommender systems have been generally classified, according to the way in which they analyze the information of the user and filter the list of items, into content-based, collaborative and hybrid systems [15].

2.1 Content based Recommendation System (CB)

Content-based (CB) systems calculate a degree of similarity between the users and the items to be recommended. The inspiration of this kind of recommendation methods comes from the fact that people have their subjective evaluations on some items in the past and will have the similar evaluations on other similar items in the future. This process is carried out by comparing the features of the item with respect to the user's preferences. So, it is assumed that both users and alternatives share a common representation (e.g., they are composed of the same set of attributes or keywords). The output of the comparison process is usually an overall performance score, which indicates the degree of matching between the user's profile and each alternative. The higher the score is, the higher the performance of the alternative for a given user. Sometimes these methods also take into account the rating history of the user.

In this approach, the recommendation system relies on having an accurate knowledge of the user's preferences to be able to select the appropriate items. This kind of approaches may suffer from the ''cold start'' problem when a new user enters in the system, because we can elicit poor knowledge about the user in an initial stage. For CB recommender systems, it is important to learn user's profiles. Various learning approaches have been applied to construct profiles of users. For example – a statistic-based approach tfidf is used to build user's profile to recommending Web pages; a reinforcement learning method was employed for Book recommendations.

2.2 Collaborative Filtering Recommendation System (CL)

Collaborative filtering systems make recommendations based on groups of users with similar preferences. The similarity between users is normally computed by comparing the ratings that they give to some of the items. When the system identifies who are the people that share similar interests with the current user, then the items that those people liked are recommended to this user. In this approach, some feedback about the provided recommendations is necessary, so as to know which items the user has liked or disliked (e.g. which places she has enjoyed visiting).

For the reason that CF methods do not require well-structured item descriptions, they are often implemented than CB methods and many collaborative systems are developed in academia and industry. There are two types of CF approach namely – user-based and item-based. The basic idea of user-based CF approach is to provide recommendation of an item for a user based on the opinions of other like-minded users on that item. The user-based CF approach initially finds out a set of nearest “neighbors" (similar users) for each user, who share similar interests or favorites. Finally, based on the ratings given by the user's “neighbors" on the item, the rating of a user on an unrated item is predicted. The basic idea of item-based CF approach is to provide a user with the recommendation of an item based on the other items with high correlations. The item-based CF approach first finds out a set of nearest “neighbors" (similar items) for each item. The item based CF try to predict a user's rating on an item based on the ratings given by the user on the neighbors of the target item.

For both user-based CF and item-based CF, to find similarity of measurement between users or items is a significant step. Pearson correlation coefficients, cosine-based similarity, vector space similarity, distance based similarity and so on are widely used as similarity measurement in CF methods.

2.3 Typicality based Collaborative Filtering approach for recommendation system

This is new approach for recommendation system in which the problems like data sparsity, accuracy in prediction are addressed. System adopts this idea from cognitive psychology in which find “neighbors" of users based on user typicality degrees in user groups (instead of the corated items of users, or common users of items, as done in traditional CF) and predict the ratings. It gives better performance, more accurate results and takes lower time cost [1].

The mechanism of typicality-based CF recommendation is as follows: First, cluster all items into several item groups. Second, form a user group corresponding to each item group (i.e., a set of users who like items of a particular item group), with all users having different degree of typicality in each of the user groups. Third, build a user-typicality matrix and measure user's similarities based on user's typicality degrees in all user groups so as to select a set of “neighbors” of each user. Then, predict the unknown rating of a user on an item based on the ratings of the “neighbors” of at user on the item.

2.4 Hybrid Recommendation System

Several recommender systems use a hybrid approach by combining collaborative and content based methods, to avoid some limitations of content-based and collaborative systems. A hybrid approach initially implement collaborative and CB methods separately and then combines their predictions by a linear combination of ratings or a voting scheme or other metrics [12].

Hybrid systems can integrate these techniques in different ways. Three approaches can be distinguished:

1. **Selection of the method:** The system incorporates DM, CB and CL methods, but only one of them is applied depending on the particular situation of each user.
2. **Sequential use**: Each recommendation technique is used in different stages of the process. Once the model has been trained, CB techniques generate the list of recommendations by computing ratings for each item based on the current and predicted values.
3. **Integrated use**: Both CB and CL techniques are combined during the execution.

For hybrid recommender systems it is also possible to combine item-based CF and user-based CF.

3. AI Techniques for Travel Package Recommendation System

This section makes a brief review of the main AI techniques and tools employed in tourism recommender systems. Many tourism recommender systems have to solve complex planning and scheduling problems, which are well known to be NP complete, and so that, it cannot be optimally solved in an efficient way [3].

3.1 Optimization Techniques

Many of the real world problems can’t be solved by the traditional mathematical techniques (linear/nonlinear programming). Most of these problems are optimization problems or decision problems. In this paper four optimization approaches are discussed as follows:

3.1.1 Ant Colony Optimization technique (ACO)

Ant Colony Optimization (ACO) is derived by the ant’s behavior for search of food. The field of ant algorithm derives models for complex problems taking inspirations from real ant’s behavior and transforming it into mathematical forms. ACO is one of the most successful techniques of the ant category algorithms.

Applicable to find out route- A set of autonomous entities (which represent the ants) cooperate through pheromone-mediated indirect and global communication to find a good solution to the travelling salesman problem. So in this case, ACO used to plan a route that goes through different points of interest around the city.
Example - The agent-based travel route recommender for **Tainan** [7] uses ACO technique.

Advantages -

1. It has power capacity to find out solutions to combinatorial optimizing problems.
2. It has advantage of distributed computing.
3. It is also easy to accommodate with other algorithms

Disadvantages –

1. Though ant colony algorithms can solve some optimization problems successfully, it’s not possible to prove its convergence.

3.1.2 Genetic Algorithm (GA)

Genetic algorithms are mostly used in searching and optimization problems. It is derived from the Darwin's theory of reproduction in nature. The GA is a method of moving from one population of chromosomes to a new population using natural selection and genetic operators. They maintain a population of structures that evolve on the basis of basic steps of GA like selection, mutation, crossover and recombination.

Applicable to construct the plan to visit a city - A genetic algorithm is used to construct the plan to visit a city. In each iteration of a cyclic process it considers a population of different possible plans, which are evaluated according to their utility for the user; the best ones are mutated and recombined via crossover to generate another population for the next iteration. After a certain number of iterations, the best plan is finally selected.
Example - **CT-Planner4**[9] uses genetic algorithm to construct the plan to visit city.

Advantages -
1. GA only requires a search range, which need only be constrained by prior knowledge of the physical properties of the system.
2. It supports multi-objective optimization.
3. Genetic algorithm is good for —noisy environments
4. Concept of genetic algorithm is easy to understand and many ways are available to speed up and improve a GA-based application.

Disadvantages -
1. The determination of the convenient parameters may be time consuming, e.g. population size, mutation rate etc.
2. We need to identify which variables are suitable to be treated as input population.

3.1.3 Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) is inspired basically by human brain functioning. Neural Networks, which are simplified models of the biological Neural Systems, is a massively parallel distributed processing system made up of highly interconnected neural computing elements that have the ability to learn and thereby acquire knowledge and make it available for use. Neural Networks are simplified imitations of the central nervous system, and therefore, have been motivated by the kind of computing performed by the human brain. A human brain's structural constituents contain neuron termed as entities, which perform computation such as pattern recognition, cognition, logical inference and so on. The approach, which has been built on simplified imitation of computing by neurons of a brain, has been called as Artificial Neural Networks [5].

Applicable to assess relevance of each context – It is possible to make recommendations adapted to the context of the user, that is composed of different factors (location, time, weather, social media sentiment and user preferences). In this case the idea of using an artificial neural network is best to assess the relevance of each context component for each user.
Example – The **VISIT system**[8] uses ANN approach to access relevance of each context.

Advantages -
1. It requires less formal statistical training.
2. It has ability to implicitly detect complex nonlinear relationships between dependent and independent variables.
3. Availability of multiple training algorithms.

Disadvantages –
1. Greater computational burden

3.1.4 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) technique is inspired by the behavior of the bird flocking. It is a technique that is used to explore the search space within a given problem for finding the parameters that are required to maximize objective of given problem [4].

The process starts by initializing a group of particles (solutions) and thereof, searches for optima by updating generations. In every iteration, best values for each particle is updated. The first one is the best solution (fitness) that has achieved so far, by an individual particle is termed as pbest. Another "best" value that is obtained so far by any particle in the population is termed as gBest. Whenever a particle takes part in the population as its topological neighbors, the best value is a local best and is termed as lbest.

Applicable to find out sparsity measure – In case of recommendation system, the sparsity measures can be calculated by various measure like Cosine, Pearson coefficient etc. The PSO approach is also used to get the better result.

Advantages-
1. PSO is efficient global optimizer for structural applications, with limited number of parameters.
2. PSO generates high quality solution with less calculation time and stable convergence.
3. It has very simple concept, easy to implement.

Disadvantages-
1. Premature convergence is major limitation of PSO.
2. Dependency on initial condition, difficulty in identifying design parameters, parametric values etc. are the problems need to be solved in PSO.

3.2 Clustering Approaches

Many tourism recommenders employ techniques based on collaborative filtering, in which the users of the system are partitioned into groups that share some common characteristics. The basic idea of these methods is that it can be appropriate to recommend to the user those items that have been positively valued by similar tourists. In any case, the automatic clustering tools developed in AI may be successfully used to classify the tourists. This section comments different alternatives that have been used in touristic recommender systems.

3.2.1 k-Nearest Neighbor (kNN)

k-Nearest Neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). It is a very simple way of associating a new user with similar past users of the system, for calculating which are the k past users of the system who were more similar to the current one. Having done that, the information on those users may be employed to provide recommendations (e.g., the activities that were more highly valued for them) [6].

In SAMAP the similarity between users is based on the preferences expressed over the concepts of a domain ontology. For instance, the system could easily infer that a user that likes Cinema is more similar to a user that enjoys Theatre than to another that prefers Sport activities [10]. Scalability is one of the main problems faced when using this method.

3.2.2 k-Means Clustering

A common option to group the users into different classes is to use the k-means algorithm. The initial seeds of the k desired clusters are established in some application-dependent way. Then there is an iterative process in which, in every step, the objects are sorted into the nearest cluster and the cluster prototypes are recalculated. The method converges when the objects belong to the same clusters in two consecutive iterations in the solution [11].

In the k-means algorithm is applied with three different purposes: to obtain a set of initial tourist segments, to obtain classes of users with similar demographic characteristics, and to classify users according to the explicit ratings they have provided.

K-means clustering is used to address the problem of scalability issue in recommendation system.

3.2.3 Fuzzy C-means Clustering

It is the uncertain version of k-means. The result of this algorithm is a fuzzy partition of a set of objects into clusters, such that each object has a degree of membership between 0 and 1 to each cluster, and the addition of the degrees of membership to all the clusters is 1.

This algorithm is both applied to users and to touristic points of interest (POIs). The system is able to derive rules that characterize them, that are used to integrate new users and new POIs to the clusters in which they fit better.

3.2.4 Super Vector Machine (SVM)

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplanes. In other words, given labeled training data (*supervised learning*), the algorithm outputs an optimal hyperplanes which categorizes new examples.

SVM is used as a classification technique in tourism recommender system. Tourist preferences on several kinds of tourist activities are stored in a vector, and the characteristics of each activity are also stored in the same way. Thus, SVMs may be used to compute the distance between the user's preferences and the recommendable items, so that the most appropriate ones can be efficiently found.

This work also proposes to build association rules, which explain the relationship between clusters of users (plus contextual information) and clusters of POIs. These rules permit to determine the kind of touristic activities that should be recommended to a certain type of users.

3.3 Management of Uncertainty

The task of recommending activities to a tourist is not simple, as there is not any clear and precise relationship between the characteristics and preferences of a visitor and the POIs available at a given destination. Some of the techniques developed in the AI field of approximate reasoning have been proposed to represent and reason about this uncertain relationship.

3.3.1 Bayesian Network

To manage this uncertainty one possible solution is to use Bayesian networks. It is an acyclic graph in which edges represent relationships of causality or influence between nodes. Nodes that do not have any parent have an associated probability table, indicating how likely they are to occur. Nodes that have n parents have a conditional probability table of 2^n nodes, indicating how likely they are to occur depending on the presence (or absence) of their parents.

A Bayesian networks is used where a number of attributes (age, nationality, occupation, income, travel motivation, etc.) influence directly on the probability that a certain touristic point is interesting for the user. Given a network in which the age, occupation and personality influence the type of user which, along with the travel motivation, influences the probability of the user liking a certain kind of touristic destinations.

3.3.2 Fuzzy Logic

Another option to manage uncertainty is the use of fuzzy logic. A fuzzy variable takes values as a series of linguistic labels. Each linguistic label has an associated fuzzy set, in which every value in the domain of reference is assigned a membership value to the set between 0 and 1 [13]. In that way, fuzzy logic provides a generalization of standard logic. Fuzzy sets and fuzzy reasoning may be used to present the preferences of the user and to calculate how they fit with the characteristics of a tourist attraction, to obtain the degree of membership of each user to different groups of users or to represent contextual aspects of the journey.

For instance, if the weather conditions are represented with a value between 0 and 1, instead of using a simple Boolean value for good/bad weather, it is possible to make a more fine grained analysis of the weather conditions and reason about its influence on the recommendation of each cultural activity.

4. Summary

Tourism recommender systems give personalized and relevant suggestions to tourists whenever they visit unknown places. Many approaches of the tourism recommender systems have been analyzed i.e. content based, collaborative, hybrid and typicality based approaches. Most of the approaches suggest points of interest in a destination according to the user preferences. There is a wide range of Artificial Intelligence techniques that are used for travel package recommender system. Optimization techniques offer cost-effective solutions to complex planning and scheduling problems of the system. Approximate reasoning techniques are applied to manage the uncertainty on the user's preferences, making them an ideal option when the user feedback is only implicit. Finally, Automatic clustering algorithms may be successfully used to classify tourists with similar tastes or similar features.

5. Acknowledgment

The authors would like to thank the publishers and researchers for making their resources available. We also thank the college authority for providing the required infrastructure and support. Finally we would like to extend our heartfelt gratitude to friends and family members.

References

[1] Y. Cai, H.-F. Leung, Q. Li, H. Min, J. Tang, and J. Li, "Typicality based collaborative filtering recommendation" Knowledge and Data Engineering, IEEE Transactions on, vol. 26, pp. 766{779, March 2014.}

[2] Q. Liu, E. Chen, H. Xiong, Y. Ge, Z. Li, and X. Wu, "A cocktail approach for travel package recommendation," Knowledge and Data Engineering, IEEE Transactions on, vol. 26,pp. 278{293, Feb 2014}

[3] J. Borrs, A. Moreno, and A. Valls, "Intelligent tourism recommender systems: A survey," Expert Systems with Applications, vol. 41, no. 16, pp. 7370 { 389, 2014.}

[4] Sagarika Bakshi, Alok Kumar Jagadev, Satchidananda Dehuri, and Gi-Nam Wang. Enhancing scalability and accuracy of recommendation systems using unsupervised learning and particle swarm optimization. Applied Soft Computing, 15(0):21 { 29, 2014.}

[5] Pawan Kumar Tiwaria, Deo Prakash Vidyarthib. "Evolutionary Computation and Its Applications: A Survey". Information Engineering Letters, ISSN: 2160-4114 Volume 2, Number 1, March, 2012.

[6] Shweta Singhal1, Shivangi Goyal2, Shubhra Goyal3 and Divya Bhatt4 A Comparative Study of a Class of Nature Inspired Algorithms, Computing For Nation Development, March 10 – 11, 2011.

[7] Lee, C. S., Chang, Y. C., Wang, M. H. (2009). Ontological recommendation multi agent for Tainan city travel. Expert Systems with Applications, 36, 6740–6753.

[8] Meehan, K., Lunney, T., Curran, K., & McCaughey, A. (2013). Context-aware intelligent recommendation system for tourism. In Proceedings of the 11th IEEE international conference on pervasive computing and communications (pp.328–331).

[9] Kurata, Y., & Hara, T. (2013). CT-planner4: Toward a more user-friendly interactive day-tour planner. In Information and communication technologies in tourism 2014 (pp. 73–86). Springer International Publishing.

[10] Martínez-Santiago, F., Ariza-López, F., Montejo-Ráez, A., & Ureña-López, A. (2012). GeOasis: A knowledge-based geo-referenced tourist assistant. Expert Systems with Applications, 39(14), 11737–11745.

[11] Gavalas, D., & Kenteris, M. (2011). A web-based pervasive recommendation system for mobile tourist guides. Personal and Ubiquituous Computing, 15, 759–770.

[12] Lucas, J., Luz, N., Moreno, M., Anacleto, R., Figueiredo, A., & Martins, C. (2013). A hybrid recommendation approach for a tourism system. Expert Systems with Applications, 40(9), 3532–3550.

[13] Pinho, J., da silva, B., Moreno, M., de Almeida, A. M., & Lopes, C. (2011). Applying recommender methodologies in Tourism sector. Highlights in practical applications of agents and multiagent systems. In Advances in Intelligent and Soft Computing (pp. 101–108). Springer.

[14] Z. Huang, H. Chen, and D. Zeng, "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering," ACM Trans. Inf. Syst., vol. 22, pp. 116{142,Jan. 2004.}

[15] G. Adomavicius and A. Tuzhilin. Towards the next generation of rec- ommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. on Data and Knowledge Engineering, 17(6):734749, 2005.

[16] Deepa Anand and Kamal K. Bharadwaj. Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities. Expert Systems with Applications, 38(5):5101 { 5109, 2011.}

[17] Shweta Tyagi and Kamal K. Bharadwaj. Enhancing collaborative filtering recommendations by utilizing multi-objective particle swarm optimization embedded association rule mining. Swarm and Evolutionary Computation, 13(0):1 { 12, 2013}.

[18] Jesus Bobadilla, Fernando Ortega, Antonio Hernando, and Javier Alcal. Improving collaborative filtering recommender system results and performance using genetic algorithms. Knowledge-Based Systems, 24(8):1310 { 1316, 2011.}