Personalized Travel Itinerary Recommendation Service Based on Collaborative Filtering and IEC

Cong Li
School of Computer Science
Sichuan Normal University
Chengdu, China
cnlicong@yahoo.cn

Jianjun Wang
Chinese Flight Test Establishment
Xi'an, China
flightwang@126.com

Li Ma
China West Normal University
Business College
Nanchong, China
cnmali@yahoo.cn

Qing Lu
School of Economics and Management
Shanghai University of Electric Power
Shanghai, China
luq1982@gmail.com

Abstract—**Personalized Travel Itinerary Recommendation Service (PTIRS) is a hot research problem in E-travel information management currently. To improve the problem, in this paper a PTIRS method based on collaborative filtering and IEC is proposed. Firstly the scale of travel itinerary sample set is reduced by collaborative filtering; secondly an IEC algorithm called IMAGA is used to accelerate convergence speed for alleviating the phenomena "user estimating tiredness"; finally the most satisfactory travel itinerary can be achieved for target user. Simulating experiment show the validity of the proposed method.**

Keywords-collaborative filtering; IEC; personalized travel itinerary; E-travel

I. INTRODUCTION *(HEADING 1)*

Personalized Travel Itinerary Recommendation Problem (PTIRP) is developed based on traditional Traveling Salesman Problem (TSP). PTIRP needs to not only modeling and solving traveler's itinerary and time, but also considering other important subjective factors such as traveler's preference on sights and his/her emotion. Due to PTIRP is relevant to mankind subjective preference, it is impossible to design an objective function for PTIRP. Hence, PTIRP is belong to a kind of complex multi-criteria decision-making problem which have no objective function and hard to be solved. Then PTIRP can be regard as tacit object decision-making problem for its analysis and solving.

However, a new traveler will meet the famous cold-start problem in the process of personalized travel itinerary recommendation because no his/her preference data or previous travel information can be used to generate recommendation result. So in Lu's early work [1], he proposed an improved IEC algorithm, which named IMAGA, for travel itinerary recommendation. In IMAGA, single itinerary is regard as an agent. The fitness estimated by user is regard as the power of agent. The living aim of agent is to increase and maximize user's degree of satisfactory, and bring to itself power. All the agents are live at an $m\times n$ (m, n $\geqslant 3$) circle-style grid, and every agent will act with other agents nearby. The area of near neighbor agents is called competition neighbor area. The scale of competition neighbor area will be different according to different problems. An agent has two actions, namely evolvement and competition, for its living and improving itself power. In the evolvement action, any agent $agent_{ij}$ will cross with the most excellent agent in its competition neighbor area by probability P_c, the most excellent agent will not be changed; $agent_{ij}$ will be replaced by next crossover generation, after the replacing process $agent_{ij}$ will mutate by probability P_m. In the competition action, if the power of $agent_{ij}$ is not lower than any agent in its competition neighbor area, then $agent_{ij}$ can be live; otherwise $agent_{ij}$ will die and be replaced by the son of the most excellent agent. More detailed algorithm description can be found in [1].

In this paper, we combine collaborative filtering algorithm [2, 3, 4] with Lu's early algorithm IMAGA, and propose a recommendation method for personalized travel itinerary. This new method can be used to help traveler plan his/her personalized travel itinerary. Simulating experiment show the validity of our new method in this paper.

II. PERSONALIZED TRAVEL ITINERARY RECOMMENDATION PROBLEM

A. Description of Tacit Object Decision-making Problem

In the decision-making problems of our production and living, there exist two kinds of decision targets: (1) Explicit Object, namely the decision targets that can be described by quantity and structure; (2) Tacit Object, namely the decision targets that can't or hard to be described by quantity and structure. Tacit object decision-making problem is the multi-objective decision-making problem, and it can be described as follows.

Based on Huang's idea in [5], in this paper tacit object decision-making problem is defined as the forming of $\langle N, X, f, c, Y, P\rangle$:

(1) N is the set of decision makers. Suppose the dimensionality of N is $|N|$, then when $|N|=1$, $|N|$ denotes single-person decision; when $|N|\geqq 2$, $|N|$ denotes group decision;

(2) X is the set of alternatives. Suppose alternative $x=(x_1, x_2, \dots, x_n)$, x_i is the decision variables which belong to universe U_i, n is the dimensionality of decision space.

(3) $f=(f_1, f_2, \dots, f_p; f_{p+1}, f_{p+2}, \dots, f_{p+q})$ is the target function. The preceding $p(p \geqslant 0)$ functions denote those target functions can be described by quantity, namely functions $f_1, f_2, \dots, f_p$ can be described as some math form of decision variables $x_1, x_2, \dots, x_n$ explicitly; the back $q(q \geqslant 1)$ functions denote those target functions can't or hard to be described by quantity. Generally tacit object decision-making problem is multi-criteria decision making problem.

(4) $c(x)=(c(x_1), c(x_2), \dots, c(x_m)) \leqslant 0$ is constraint condition. It is relevant to the requests of problem.

(5) $Y=\{(y_1, y_2, \dots, y_p, \dots, y_{p+q})\}$ is the set X $(c(x)\leqslant 0)$ under constraint condition, namely $Y=\{y \mid y=f(x), c(x)\leqslant 0, x \in X\}$. The total targets of optimization are as follows:

$$\text{Maximize } y = f(x)$$

$$\text{subject to } c(x) \leqslant 0$$

(6) P is the decision makers' "accept or reject" criteria. Due to the relativity between decision criteria of tacit object decision-making problem and person, so different person has various decision criteria. Then the target function f will be different forms. The decision criteria will be more complex under multi-person.

B. Description of Personalized Travel Itinerary Recommendation Problem

Personalized Travel Itinerary Recommendation Problem is a tacit object decision-making problem, and it accords with the three characteristics of tacit object decision-making problem: the decision targets hard to be described by quantity and structure; the decision makers' preference will be changed by the process of decision analysis; usually the decision problems are NP problem. The basic concepts and description of PTIRP are given as follows.

Basic concepts:

(1) Set of sights: $sight = \{s_i \mid i = 1, 2, \cdots, n\}$, i denotes the ith sight in a travel area; n denotes the count of all sights in the travel area.

(2) Itinerary segment: the itinerary between any two sights.

(3) Itinerary T: the rank of sights, it denotes the access sequence of sights. $T = (s_{i1}, s_{i2}, \cdots, s_{in})$, $(i1, i2, \cdots, in)$ is a rank of $(1, 2, \cdots, n)$.

(4) Cell attributes: the information about sights such as entrance ticket fee and the time for arriving sights.

(5) Double attributes: the information on itinerary segment, for example, the distance and time between two sights.

(6) Add-enabled attributes: the attributes which can be accumulated in cell attributes and double attributes, for example, the entrance ticket fee and fare can be accumulated to be the total itinerary fee.

(7) Add-unable attributes: the attributes which can't be accumulated in cell attributes and double attributes, for example, the time for arriving every sight.

(8) Itinerary attributes: the information showed by travel itinerary, and it is the main attribute concerned by traveler, such as the total fee, total time and some add-unable attributes concerned by traveler.

Problem description:

Personalized travel itinerary means that traveler should choose an appropriate itinerary to access every sights he/she wants to access in a travel area, and the time (time point, time cost), the fee and the personal easiness (such as the sight suits the traveler's preference, the distance he/she needs to walk) will bring the biggest satisfaction to himself/herself. Hence, the task of personalized travel itinerary recommendation problem is to find out the travel itinerary. It can be described as follows:

$$S(u) = \operatorname{argmax}(T) \tag{1}$$

In (1), $T \in N$, N is the set of all the itineraries; $S(u)$ is traveler u's degree of satisfaction. Owing to the factors which affect traveler's satisfaction are not only the targets can be quantized (distance, time and fee of itinerary) but also the targets can not or hard to be quantized (traveler's preference of arriving sights), argmax (T) can't be described by function. It is necessary to converge the problem by using Interactive Genetic Algorithm (IGA) for the direction of traveler's satisfactory itinerary.

III. Recommendation Service Based on Collaborative Filtering and IEC

A. User-based Collaborative Filtering Algorithm

1) Representation

A nearest-neighbor algorithm uses a user-item ratings matrix $R(m,n)$ to represent users' information. $R(m,n)$ is a $m \times n$ matrix, in which m rows denote m users, n columns denote n items, and $R_{i,j}$ is user i's rating on item j.

2) Neighborhood formation

As to a target user u, the algorithm needs to find out u's nearest neighborhood $U = \{u_1, u_2, \cdots, u_K\}$, $u \notin U$. The similarity $sim(u,u_k)(1 \le k \le K)$ between u and another user u_k is calculated in a descending order. Some common user similarity measures are showed in Table 1. The value of k

can be determined by a given constant or similarity threshold. The aforementioned methods can be combined to determine the value of k, including the choice of the top k users that have the highest similarity than the given threshold.

In Table Ⅰ [6], u and v represent two random users with $u \neq v$; $sim(u,v)$ denotes the similarity between u and v; I_{uv} denotes the common ratings set between u and v in the sense that for $\forall i \in I_{uv}$, there exists $R_{u,i} \neq \varnothing$ and $R_{v,i} \neq \varnothing$; $\vec{u}$ and $\vec{v}$ represent u 's and v 's rating vectors on I_{uv} respectively; $R_{u,i}$ and $R_{v,i}$ represent u 's and v 's ratings on item i respectively; $\overline{R_u}$ and $\overline{R_v}$ represent u 's and v 's average ratings on I_{uv} respectively; R_{med} represents the median used in recommender system.

TABLE I. USER SIMILARITY MEASURES

Measure name	Measure description
cosine similarity	$sim(u,v) = \cos(\vec{u},\vec{v}) = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|_2 \times \|\vec{v}\|_2} = \frac{\sum_{i\in I_{uv}} R_{u,i}\cdot R_{v,i}}{\sqrt{\sum_{i\in I_{uv}} R_{u,i}^2}\sqrt{\sum_{i\in I_{uv}} R_{v,i}^2}}$
Pearson correlation coefficient	$sim(u,v) = \frac{\sum_{i\in I_{uv}} (R_{u,i}-\overline{R_u})\cdot(R_{v,i}-\overline{R_v})}{\sqrt{\sum_{i\in I_{uv}} (R_{u,i}-\overline{R_u})^2}\sqrt{\sum_{i\in I_{uv}} (R_{v,i}-\overline{R_v})^2}}$
constrained Pearson correlation coefficient	$sim(u,v) = \frac{\sum_{i\in I_{uv}} (R_{u,i}-R_{med})\cdot(R_{v,i}-R_{med})}{\sqrt{\sum_{i\in I_{uv}} (R_{u,i}-R_{med})^2}\sqrt{\sum_{i\in I_{uv}} (R_{v,i}-R_{med})^2}}$

3) Recommendation generation

The generation methods of top-N item recommendation set for target users include predicting rating-based recommendation, association rule-based recommendation and most-frequent item recommendation. In the aforementioned methods, predicting rating-based recommendation is most widely used. In this method, if u 's neighborhood $U = \{u_1, u_2, \cdots, u_K\}$ is generated, the rating item sets of users in U are $I_1, I_2, \cdots, I_K$ respectively and u 's rating item set is I_u with $I_w = I_1 \cup I_2 \cup \cdots \cup I_K - I_u$, then for $\forall i \in I_w$ there exists $R_{u,i} = \varnothing$, so u 's rating on i can be predicted by $P_{u,i}$ as follows:

$$P_{u,i} = \overline{R_u} + \frac{\sum_{(u_k \in U)\cap(R_{u_k} \neq \varnothing)} sim(u,u_k) \times (R_{u_k,i} - \overline{R_{u_k}})}{\sum_{(u_k \in U)\cap(R_{u_k} \neq \varnothing)} (|sim(u,u_k)|)} \qquad (2)$$

In (1), $R_{u_k,i}$ stands for u_k 's nonempty rating on item i, $\overline{R_{u_k}}$ for the average rating on the common rating item set between u_k and u [7], and $\overline{R_u}$ for u 's average rating on all items. Then the top-N items can be selected to form a top-N item recommendation set $I_{rec} = \{i_1, i_2, \cdots, i_N\}$ by a descending order on $P_{u,i}$.

B. Recommendation Process of Personalized Travel Itinerary and Its Realization

The basic process of personalized travel itinerary recommendation is as follows:

Firstly, collaborative filtering is used to search and find out active traveler's nearest neighborhood set. All the travelers in the set will have the same or similar travel interests, including sights preference and travel itineraries.

Secondly, let active traveler clicks the travel itinerary figures of his nearest neighbors on the screen, then system will evolve and close to active traveler's satisfactory itinerary by using IEC algorithm until active traveler stop his evolve-clicking action.

The process of our proposed method can be showed by Fig. 1 as follows:

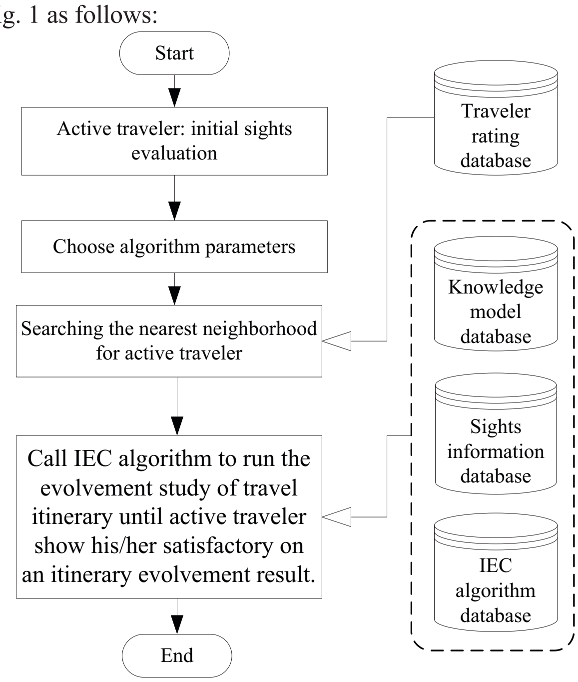

Figure 1. Recommendation process of the proposed method.

We take "one day travel in Hefei" as the background of application research, and realize the program of personalized travel itinerary recommendation based on IMAGA algorithm. The choose itinerary attributes include total distance, total fee, itinerary time, total time, count of joining activities, the time of arriving every sight. According with [1], the population size of IMAGA is 12 and ranged by three rows and four columns, crossover probability P_c =0.6, mutation probability P_m =0.1, λ =0.3. The termination manner of the algorithm is manual termination manner by active traveler's choosing on his/her satisfactory itinerary. Fig. 2 shows the using of collaborative filtering in E-commerce personalized recommendation (after few changing, collaborative filtering algorithm can be used in E-travel personalized itinerary recommendation), and Fig. 3 shows the evolving of personalized travel itineraries based on IMAGA.

Figure 2. The using of collaborative filtering in E-commerce.

Figure 3. Evolving the personalized travel itineraries based on IMAGA.

The results of simulating experiment show that the new recommendation method in this paper has the following advantages:

(1) The scale of travel itineraries set is reduced by collaborative filtering, which favors the realization of whole converging of algorithm and alleviates the user tiredness;

(2) Diversity of population will be retained by IEC algorithm, and in the evolvement process new agents can inherit the excellent characteristics of near neighbor agents, thus also can improve the whole converging speed of algorithm.

IV. CONCLUSION

Personalized Travel Itinerary Recommendation Service (PTIRS) is a hot research problem in E-travel information management currently. In PTIRS, when a new traveler comes to some travel websites, he/her will meet the cold-start problem while his/her personalized travel itinerary recommendation is send to him/her. Hence, the PTIRS method based on collaborative filtering and IEC is proposed in this paper. By using collaborative filtering we can reduce the scale of travel itinerary sample set. Moreover we use the IEC algorithm, IMAGA, to accelerate convergence speed for alleviating the phenomena "user estimating tiredness". So, the traveler can obtain his/her most satisfactory travel itinerary. The next work is to improve the performance of IMAGA and collaborative filtering algorithm, and reduce the user time spending on evolving of travel itineraries.

ACKNOWLEDGMENT

This work was supported by the Scientific Research Fund of SiChuan Provincial Education Department (No. 09ZB068), the Scientific Research Fund of Sichuan Normal University (No. 037185). We would like to thank anonymous reviewers for their valuable comments.

REFERENCES

[1] Qing Lu, Chang-yong Liang, Yong-qing Huang and Jun-ling Zhang, "Interactive multi-agent genetic algorithm for travel itinerary planning," Application Research of Computers, vol. 25, Nov. 2008, pp. 3311-3313, doi: CNKI:SUN:JSYJ.0.2008-11-034.

[2] J. Herlocker, J. A. Konstan, A. Borchers and J. Riedl, "An algorithmic framework for performing collaborative filtering," Proc. of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieve, ACM Press, 1999, pp. 230-237, doi: 10.1145/312624.312682.

[3] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Analysis of recommendation algorithms for E-commerce," Proc. of the 2nd ACM Conference on Electronic Commerce, ACM Press, 2000, pp. 158-167, doi: 10.1145/352871.352887.

[4] J. Herlocker, J. A. Konstan and J. Riedl, "An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms," Information Retrieval, vol. 5, Oct. 2002, pp: 287-310, doi: 10.1023/A:1020443909834.

[5] Yong-qing Huang, "Research on decision process and decision methods of decision-making problem with tacit objective function based on IEC," Ph.D. dissertation, Hefei: Hefei University of Technology, China, 2006.

[6] H. J. Ahn, "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem," Information Sciences, vol. 178, 2008, pp. 37-51, doi: 10.1016/j.ins.2007.07.024.

[7] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "Grouplens: an open architecture for collaborative filtering of netnews," Proc. of the 1994 ACM on Computer Supported Cooperative Work, ACM Press, 1994, pp. 175-186, doi: 10.1145/192844.192905.