HOTELS RECOMMENDATION METHODS BASED ON PATTERN RECOGNITION AND FACTORIZATION MACHINES

Y. CHEN † and D. YAO‡

† *School of Computer Science and Engineering, Yulin Normal University, Yulin, China.*
709418311@qq.com
‡ *School of Computer Science and Engineering, Huaihua University, Huaihua, China.*
hhxyydh@163.com

***Abstract*— The traditional recommendation methods for hotels usually compute rating similarity and make recommendation based on collaborative filtering algorithm. It is due to having no consideration of the tourist' and hotels' multi-faceted attributes. Thus, the accuracy of recommendation would be affected. To solve this problem, a series of formal methods are adopted to define the various attributes of hotels and tourists. To begin with it, get hotel star factor, hotel hardware facilities, cost performance, geographical location and the characters of the preference of the star of the tourists, and after that a partial weighting model is used to compute a recommended label value. Finally, the factorization machines (FMs) is used to make recommendations. The experimental results show that the proposed methods can solve data sparseness problem to some extent. Additionally, both its recommendation and ranking accuracy are better than those of the traditional collaborative filtering algorithm, which can improve the tourist satisfaction in personalized hotels recommendation.**

***Keywords*— factorization machines; highly sparse data; normalized treatment; pattern recognition.**

I. INTRODUCTION

With the advent of the DT (Data Technology) era, big data has drawn more and more attention. Big data is not "big" but "useful" (Wu, 2015). In the face of mass data and information, how to process these data with huge commercial value and eventually transform the commercial value of data to many industries to win the competition is the key. Therefore, big data has become a hot research topic nowadays.

The recommender system is a product that has come into being and has tremendous potential to easily liberate users from the heap of information. It is based on the user's historical behaviors to analyze the speculation of the user's needs and interests. It is an advanced business intelligence platform for users to provide personalized decision support and information services to users (Wu, 2018a).

Recommended systems commonly have collaborative filtering recommendation, recommendation based on content, recommendation based on association rules, recommendation based on utility, recommendation based on knowledge, combined recommendation, and so on. In the aspect of hotel recommendation, Fu-guo (2014) solved the cold start problem and scalability problem of traditional collaborative filtering using clustering-based collaborative filtering and Rankboost algorithm. Sun (2014) proposed a hotel recommendation model based on user behavior characteristics, which can effectively improve the quality of recommendation. Ke (2018) reveals the granulation monotonicity and systematic relationships of three-way weighted combination-entropies. Wu *et al.* (2018b) presented a methodology for control structure design based on the Singular Value Decomposition (SVD) of the Hankel matrix. At the same time, user-item interaction factor is added, improving the novelty and diversity of the recommendation list.

II. BASIC CONCEPTS

The hotel recommendation system can be defined as follows: let T be the set of all tourists and the collection of tourist hotels in all online regions as H. Suppose the utility function *f*() is used to calculate the recommendation *R* of online hotel *h* for tourists *t*, that is, $f: T \times H \to R$. Find the online hotel set H^* with the greatest recommendation degree *R*, which is a personalized travel hotel to be recommended for tourists *t*, that is, $\forall t \in T, H^* = \arg \max_{h \in H} f(t,h)$ 。

Factorization machine model: Steffen is a factorization model that integrates a factorial matrix with matrix factorization. It decomposes the interaction from category variables by factorization, and removes the autocorrelation terms from the polynomial regression. It is particularly suitable for the highly sparse data set that gives predictions for the true value eigenvectors.

In the current application researches, the second-order factorization machine is a common model. Assuming that the input eigenvector is *x* and the target variable is *y*, the predicted value $\hat{y}$ of *y* is given by Eq. 1:

$$\hat{y}(x) = v_0 + \sum_{i=1}^{m} v_i x_i + \sum_{i=1}^{m-1} \sum_{j>i}^{m} \langle u_i, u_j \rangle x_i x_j \tag{1}$$

In $v_0 \in \Re$, $v \in \Re^m$, $u \in \Re^{mxk}$ and factorization dimension $k \in \Re^+$ $(k \in m)$, let u_i denotes the *i* the variable of *m* attribute factors. First eliminate the autocorrelation terms, which define *j*>*i*, then factorize the interaction in other types of variables. Considering only the interaction $\langle u_i, u_j \rangle$ between the *i* th and *j* th disparate eigenvalues, we approximate it with the inner product of two low-rank matrices:

$$v_{i,j} \approx \langle u_i, u_j \rangle = \sum_{f=1}^{k} u_{i,f} u_{j,f} \text{ , } \quad v_{i,j*} \approx \langle u_i, u_{j*} \rangle = \sum_{f=1}^{k} u_{i,f} u_{j*,f} \quad (2)$$

The high-order eigenvectors represent the interactions between features and project them into an implicit space. The factorization machine can contain multiple class variables and is suitable for highly sparse data backgrounds, because $v_{i,j}$, $v_{i,j*}$ share some kind of interaction.

The reason why the factoring model is particularly suitable for datasets of high sparseness and owning multiple categories of variables is that they share a certain role among multiple categories of variables during factorization. The factorization machine is equivalent to the traditional matrix decomposition model only if there are 2 category variables. For example, in users and product data sets, the characteristic variables are users (User={*A*,*B*,*C*}) and commodities (Commodity={*Phone*,*Computer*,*Bicycle*,*Camera*}).

This time, the estimation formula of the score should be:

$$\hat{y}(x) = t_0 + t_u + t_i + \langle v_u, v_i \rangle \quad (9)$$

If there are more than two feature variables in the dataset, the feature vectors are all combined into two groups and then expanded them as Formula 1 according to the number of features required. As the example above with the product label={*L1*,*L2*,*L3*,*L4*}） eigenvectors, the score estimation formula expands directly to:

$$\hat{y}(x) = t_0 + t_u + t_i + \langle v_u, v_l \rangle + \langle v_i, v_l \rangle + \langle v_u, v_i \rangle \quad (10)$$

III. DATA PREPROCESSING

For the sake of the smooth use of the factorization machine model for regional tourism hotel recommendation, the relevant non-consistent raw collected data must be preprocessed. According to the actual recommended situation for the tourist hotel, the following formal definitions are used to normalize the data for the purpose of expediting the training and prediction of the model and to improve the recommendation accuracy of tourist hotels.

Definition 1：Hotel star factor (*Hsf*)

If we know that $0 < Hsf \le 1$， we will know the normalized value of the hotel star, which indicates that the larger the value presents the higher the star rating of the hotel, which is given by:

$$Hsf = \frac{Hs}{5} \quad (11)$$

Hsf means the star of the hotel.

Hence, $Hs \in \{1,2,3,4,5\}$ 。

Definition 2：Hotel old and new coefficient (*Honc*)

Knowing that $0 < Honc \le 1$, we have that the *Honc* is used to quantify the hotel old and new. The greater the coefficient means the new hotel, which explained as follows:

$$Honc = \left(\frac{1}{Yc} + \frac{1}{Yr}\right) \Big/ 2 \quad (12)$$

Yc = *Curyear* – *Build* + 1 indicates the current year minus the number of years and plus 1 (avoiding denominator to be zero).

Yr = *Curyear* – *Revision* + 1 denotes the current year minus the renovation then plus one.

Definition 3：Hotel scale factor (*Hscf)*

Let us consider that $0 < Hscf \le 1$， then, the *Hsf* is used to quantify the size of the hotel, mainly to measure the number of rooms. A *Hscf* greater means the more hotel rooms and the larger the scale, which is reflected by the following formula:

$$Hscf = \frac{R}{R_{\max}} \quad (13)$$

In the above formula, *R* represents the number of the hotel rooms. $R_{\max}$ represents the maximum number of all online hotel rooms.

Definition 4：Hotel Facilities Service Coefficient (*Hfsc*)

In this case, we have that $0 < Hfsc \le 1$ and here the *Hfsc* is used to quantify the main facilities of the hotel, such as its major services and the room ones, including the completeness of hardware and software. The greater coefficient reflects the more complete facilities and the more considerate services, which is explained by:

$$Hfsc = \left(\frac{Mf}{Mf_{\max}} + \frac{Ms}{Ms_{\max}} + \frac{Hf}{Hf_{\max}} + \frac{Rf}{Rf_{\max}}\right) \Big/ 4 \quad (14)$$

Here, *Mf*, *Ms*, *Hf*, and *Rf* represent the number of major facilities, main services, hotel facilities and room facilities of the current hotels respectively.

$Mf_{\max}$, $Ms_{\max}$, $Hf_{\max}$ and $Rf_{\max}$ indicate the largest number of major facilities in all hotels and the number of major services , hotel facilities and the number of room facilities.

Definition 5：Low price ratio (*Lpr*)

When $0 \le Lpr < 1$, the *Lpr* is used to quantify the price of the hotel compared with the same regional star hotel. Thus, the greater value shows the more dominant price, which is expressed by the following formula:

$$Lpr = 1 - \frac{p}{p_{\max}} \quad (15)$$

At this time, *p* is the lowest online booking price of the tourist hotel and $p_{\max}$ is the maximum of the lowest prices of all online hotels in the same area.

Definition 6：Scenic concentration index (*Sci*)

If we have that $0 < Sci \le 1$， then the *Sci* is used to quantify the concentration of the hotel near the area. The greater the value expresses the more concentrated area around the hotel, or the surrounding area. So, there is the higher the average value of stars, the higher the possibility for visitors to choose the hotel to stay. It explains as follows:

$$Sci = \frac{2 \times \sum_{i=1}^{n} \frac{1}{1 + dis_i} \times \frac{star_i}{5}}{n} \quad (16)$$

In this place, *dis* says the distance between the hotel and the area, *star* says the area star, and *n* says the number of scenic spots near the hotel, besides the text taken *n* = 10 here.

Definition 7: Transport facilitation index (*Tfi)*

Stating that $0 < Tfi \le 1$, then the *Tfi* is used to quantify the convenience of hotels for going to airports, railway stations and high-speed rail stations. The greater the value shows the more convenient traffic between the hotels and these three transportations, which is expressed by the following formula:

$$Tfi = \frac{\sum_{i=1}^{n} \frac{1}{1+dis_i}}{n} \tag{17}$$

Here, *dis* says the distance between hotels and the airports, train stations, high-speed rail stations, *n* says the number of airports, railway stations, high-speed rail stations. In the text, *n* = 3.

Definition 8: Hotel praise rate (*Hpr)*

Let $0 \le Hpr \le 1$. The *Hpr* is used to quantify online bookings of tourist hotels for tourists of the positive review ratings. The greater the value indicates that tourists are more satisfied with the hotel, which is showed through the formula as below:

$$Hpr = \frac{n_{Pr}}{n_{Gc}} \tag{18}$$

n_{Pr} is the number of praise in a tourist hotel.

n_{Gc} is its total rating.

Definition 9: Comment factor (*Cf)*

Let $0.1 \le Cf \le 1$. Here *Cf* is used to quantify the normalized value of a general comment of the tourists on an online travel hotel. The larger the value indicates the more the hotel occupancy, the more popularity. As specified in the following Table 1, rules of normalization give a comment factor *Cf*:

Table 1. Comment factor normalization rule.

Comments *n*	Factor *Cf*	Comments *n*	Factor *Cf*
[0,300)	0.1	[1500,1800)	0.6
[300,600)	0.2	[1800,2100)	0.7
[600,900)	0.3	[2100,2400)	0.8
[900,1200)	0.4	[2400,3000)	0.9
[1200,1500)	0.5	⩾3000	1.0

Definition 10: Tourist star preferences (*Tsp*)

Having established that $0.1 \le Tsp \le 1$. The *Tsp* is used to quantify the tourist preference for tourist hotel stars. The larger the value indicates that tourists prefer the more upscale hotels. *Tsp* is normalized using the min-max normalization, $Tsp \in [T_{\min}, T_{\max}]$ ($T_{\min}$ = 0.1 and $T_{\max}$ = 1.0. In the text, they are used to represent the minimum preference value and the maximum preference value respectively) whose value is given by:

$$Tsp = \frac{Th - Th_{\min}}{Th_{\max} - Th_{\min}} \times (T_{\max} - T_{\min}) + T_{\min} \tag{19}$$

$Th_{\min}$ and $Th_{\max}$ represent the lowest and highest preference of all the tourists respectively. *Th* represents the average value of a star-rated hotel where a tourist stays. That is, the total number of hotel stars and occupancy times divided by total occupancy times, which is showed y by the following expression:

$$Th = \sum_{i=1}^{5} Hs_i \times n_i \Big/ \sum_{i=1}^{5} n_i, n_i \in \mathbb{Z}^+ \tag{20}$$

In the above formula, Hs_i indicates the star number of the tourist hotels to stay, and n_i indicates the number times of checking in Hs_i.

According to the above definition of the raw collected data for pretreatment, it can be obtained as shown in Table 2 and Table 3 of the normalized data on hotels and tourists:

Table 2. Hotel property normalized data.

Hotel ID	*Hsf*	*Honc*	*Hscf*	*Hfsc*	*Lpr*	*Sci*	*Tfi*	*Hpr*	*Cf*
Ht001	1.0	0.0836	0.8859	0.8459	0.7433	0.0946	0.1727	0.7024	0.9
Ht002	0.6	0.0714	0.1033	0.7294	0.9339	0.0845	0.2078	0.6316	0.1
Ht003	0.8	0.0833	0.5217	0.7129	0.8430	0.0843	0.1915	0.6757	0.1
Ht004	0.4	0.3000	0.0408	0.5434	0.9731	0.1005	0.1658	0.7184	0.1
Ht005	0.4	0.2083	0.0652	0.6264	0.9271	0.0932	0.1756	0.7909	0.1
Ht006	0.8	0.1455	0.1495	0.7175	0.9283	0.0863	0.1939	0.6946	0.7

Table 3. Tourist star preference normalized data.

Visitors ID	N_5	N_4	N_3	N_2	N_1	*Th*	*Tsp*
Vis001	3	4	2	0	0	4.1111	0.7818
Vis002	6	21	50	64	19	2.5688	0.4032
Vis003	5	3	4	6	1	3.2632	0.5737
Vis004	4	4	3	3	5	2.9474	0.4962
Vis005	6	0	0	0	0	5.0000	1.0000
Vis006	12	5	10	4	4	3.4857	0.6283

IV. FM MODEL DESIGN

Not only regional tourist hotel is a large number set, but also tourists are a large number of sets. However, tourists only stay in the limited hotel travel. In this way, the experimental data which sets up with hotel and tourist related factors, most of the value elements will be 0. In other words, the data set is highly sparse. In this case, in order to accurately recommend tour hotels for visitors quickly and efficiently, it adopted hotel star factor (*Hsf*), hotel hardware facilities (*Hhf*), low price ratio (*Lpr*), geographic location (*Gl*), tourists evaluation (*Te*), tourist star preference (*Tsp*) as the eigenvectors $X=\{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\}$ in the FMs model. The total number of vectors correspond to the eigenvalues of the element $x^{(i)}$, where *Hhf*, *Gl* and *Te* are given by definitions 11-13 respectively.

Definition 11: Hotel hardware facilities (*Hhf*)

We have that $0 < Hhf \le 1$, so *Hhf* is used to quantify the aggregate objective factual value of hardware facilities, such as the level of new and old hotels, the size of facilities and the facilities and services. The larger the value means the newer the hotel, the more completed hardware facilities and the more thoughtful service. Its value is given by:

$$Hhf = (Honc + Hscf + Hfsc) / 3 \tag{21}$$

Definition 12: Geographic location (*Gl*)

Taking into account that $0 < Gl \le 1$, then the *Gl* quantifies the convenience of the location of the hotel, including the distances from the hotels to the scenic spots and the distance from the hotels to the airport stations. The larger the value of *Gl* indicates the more convenient the location of the hotel, which is expressed by:

$$Gl = \lambda_1 \times Sci + \lambda_2 \times Tfi \tag{22}$$

λ_1 and λ_2 are biased weights, which respectively represent the specific gravity of *Sci* and *Tfi* in *Gl*. The value of λ1 and λ2 can be set by tourists according to their needs, and $\lambda_1 + \lambda_2 = 1$.

Definition 13： Tourist evaluation (*Te*)

Let $0 \le Te \le 1$. Hence, the *Te* is used to quantify online transactions of tourist hotels and the satisfaction of the tourists with tourist hotels. The larger the *Te* value illustrates the greater the number of online transactions of the tourist hotels, and the higher satisfaction of the tourists. It can be explained as below:

$$Te = n_1 \times Hpr + n_2 \times Cf \tag{23}$$

This article may as well suppose n_1=0.6，n_2=0.4。 After that, $y^{(i)}$ is calculated as the recommended label value of the FMs model according to the recommended bias with the emphasis on formula (24), and all $y^{(i)}$ form the target vectors $Y=\{y^{(1)},\ y^{(2)},\ \ldots,\ y^{(n)}\}$.

$$y^{(i)} = \varphi_1 Hsf_i + \varphi_2 Hhf_i + \varphi_3 Lpr_i + \varphi_4 Gl_i + \varphi_5 Te_i + \varphi_6 Tsp_i, \sum_{m=1}^{6} \varphi_m = 1 \tag{24}$$

In the above formula, Hsf_i, Hhf_i, Lpr_i, Gl_i, Te_i, Tsp_i are the values of the variable $x^{(i)}$ on the six features respectively. Making a recommendation should be based on the evidences covering hotel star factors, the hotel hardware facilities, cost performance, locations, tourist evaluation and tourist star preferences and other factors.

Among them, φ_1, φ_2, φ_3, φ_4, φ_5 and φ_6 are the corresponding weighted values. In the recommendation, the corresponding biased weight can be set according to the different bias, and the recommended value $y^{(i)}$ can be obtained with different emphasis. A large weight coefficient indicates that the recommendation is biased toward this feature variable. For example, if φ_4 is set to be the maximum, it indicates that it is recommended to focus on the location of the hotel. This set of biased weights is convenient for tourists to give more accurate recommendation results according to their different requirements.

Finally, *X* and *Y* form an example of the FMs prediction model showing in Fig. 1:

	Feature Vector *X*														Target *Y*	
	Tourist			Hotel					*Hsf*	*Hhf*	*Lpr*	*Gl*	*Te*	*Tsp*		
$x^{(1)}$	1	0	0	1	0	0	0	0	1.0	0.6051	0.7433	0.1258	0.7814	0.7818	0.5676	$y^{(1)}$
$x^{(2)}$	1	0	0	0	1	0	0	0	0.6	0.3014	0.9339	0.1338	0.4454	0.7818	0.4547	$y^{(2)}$
$x^{(3)}$	1	0	0	0	0	0	1	0	0.8	0.4393	0.8430	0.1272	0.4454	0.7818	0.4912	$y^{(3)}$
$x^{(4)}$	0	1	0	1	0	0	0	0	1.0	0.6051	0.7433	0.1258	0.7814	0.5909	0.5294	$y^{(4)}$
$x^{(5)}$	0	1	0	0	0	1	0	0	0.4	0.2947	0.9731	0.1266	0.4710	0.5909	0.3995	$y^{(5)}$
$x^{(6)}$	0	1	0	0	0	0	1	0	0.8	0.4393	0.8430	0.1272	0.4454	0.5909	0.4530	$y^{(6)}$
$x^{(7)}$	0	0	1	0	1	0	0	0	0.6	0.3014	0.9339	0.1338	0.4454	0.3864	0.3756	$y^{(7)}$
$x^{(8)}$	0	0	1	0	0	1	0	0	0.4	0.2947	0.9731	0.1266	0.4710	0.3864	0.3586	$y^{(8)}$
$x^{(9)}$	0	0	1	0	0	0	0	1	0.4	0.3000	0.9271	0.1262	0.5145	0.3864	0.3593	$y^{(9)}$
	Vis001	Vis002	Vis003	Ht001	Ht002	Ht003	Ht004	Ht005								

Figure 1. Examples of FMs prediction models.

Each row in Fig. 1 represents a feature vector $x^{(i)}$ and its corresponding Target $y^{(i)}$, which shows an example of how to extract feature vectors from the data (taking tourist hotels and tourists as an example).

For example, hotel star preference Tsp of the tourist named Vis001 is 0.7818 and he or she stayed in Guilin, Guangxi tourist hotels called Ht001, Ht002, Ht004; In tourist hotel named Ht001, its *Hsf*, *Hhf*, *Lpr*, *Gl*, *Te* are 1.0, 0.6051, 0.7433, 0.1258, 0.7814.

V. EXPERIMENT AND RESULT ANALYSIS

The data comes from the data of Guilin's tourist hotels in the recent two years on the official website called '*Mafengwo*', including 12 popular accommodation areas, 2 surrounding areas, 3,542 online hotels, 946425 tourists and 1,479,783 reviews. From these data, we can see that the average number of tourist reviews about hotels is about 418, and the number of reviews per tourist is less than 2.

In order to measure the evaluation index of the accuracy of each recommended experimental result, this paper refers to the evaluation index (Yu, 2018) by using the mean absolute error (MAE), and defines the evaluation index as follows:

$$MAE = \sum_{th \in T_S} \left| R_{th} - R_{th}^{*} \right| \Big/ \left| T_S \right| \tag{25}$$

Where $|T_S|$ is the size of test set T_S, R_{th} is the true degree of recommendation of tourist *t* for tourist hotel *h* in the test set, R_{th}^{*} is the predicted value of tourist *t* for tourist hotel *h*.

To make the experiment plausible, each experiment took 20% of each tourist and hotel in the experimental data set (*E*) as the test set (T_E) and the remaining 80% of the experimental data (E-T_E) as the training set to make recommending Hotel experiments for tourists many times. Then the average MAE value of each experiment is taken as the final MAE value. The experimental design is as follows:

Experiment 1: Comparison of recommended precision on various optimization algorithms of FMs

A set of experimental data sets divided according to the above requirements used SGD, ALS and MCMC three algorithms to recommend Hotel for tourists in experiments. Owing to high sparseness of experimental data set E, it sets the FMs model parameter hyper-value *k* is a small number (respectively 12, 12, 20, 24, 28, 32, 36, 40). In order to speed up the convergence while setting the initialized standard deviation parameter called '*init_stdev*' to be 0.01 and set the learning rate parameter of SGD = 0.01 [7]. After the experiment, the comparison of the recommended accuracy shown in Fig. 2:

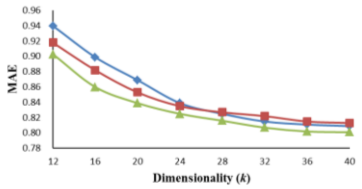

Figure 2. Accuracy comparison.

As can be seen from the figure above, all three optimization algorithms under the highly sparse experimental dataset can correctly recommend the hotel, and the prediction accuracy can meet the requirements of the recommended accuracy. Among them, the recommendation accuracy of SGD and ALS algorithms is weaker than that of MCMC algorithm. When the value of dimension hyper-parameter k reaches about 24, the SGD algorithm begins to outperform the ALS algorithm.

Experiment 2: FMs model three algorithms and user-based collaborative filtering algorithm recommended accuracy comparison.

User-based collaborative filtering recommendation algorithm, whose idea is that if some tourists graded a hotel with a similar value, you can refer to these similar values from visitors about the hotel's rating to estimate the target hotels which is unrated for tourists. It uses data statistics to find a set of neighbors with similar interests and preferences of the target tourists, and then based on the score of the nearest tourists to a specific hotel, a certain model method is used to predict the score from the target tourists. Finally, recommend the top N hotels with the highest forecast score to visitors.

In order to obtain the similarity of tourists, let I_{ij} be the set of hotels evaluated jointly by tourists *i* and tourists *j*. The similarity between them named *sim* (*i*, *j*) is calculated by Pearson's correlation coefficient formula:

$$sim(i,\mathrm{j}) = \sum_{c\in I_{ij}}(r_{ic}-\bar{r}_i)(r_{jc}-\bar{r}_j) \Big/ \sqrt{\sum_{c\in I_{ij}}(r_{ic}-\bar{r}_i)^2}\sqrt{\sum_{c\in I_{ij}}(r_{jc}-\bar{r}_j)^2} \quad (26)$$

$\bar{r}_i$ and $\bar{r}_j$ are the average rating of the hotel from both tourists *i* and *j* respectively, and r_{ic} and r_{jc} respectively represent the rating of the hotel *C*.

Then, let *X* be the set of the nearest neighbor tourists of *x*, then the predicted score P_{xi} of *x* for the non-rated hotel *i* may be obtained from the score of the nearest neighbor visitors for the hotel. Furthermore, according to Formula 27, choose the top N hotels after calculating as the recommended results.

$$P_{xi} = \sum_{k\in X} sim(x,k)(r_{ki}-\bar{r}_k) \Big/ \sum_{k\in X}|sim(x,k)| \quad (27)$$

In order to compare the recommendation accuracy of FMs model and traditional collaborative filtering model on different sparseness data sets, in the experimental data set *E*, the number of the scores of the tourists is divided into 10 tourist sets or group according to the scores of the tourists for the number of hotels. Each group of tourist is composed of those who grades for the hotels (its number $\leq 10 \times n$ (n = 1, 2, ..., 10). The latter set of tourists includes the previous set of tourists. As the number of tourists increases, the scores of the hotels also increase. The degree is also on the rise, because with the increase in the number of rating hotels, tourists only have ratings on some hotels, which will increase the sparsity of data sets.

In the experiment, FMs model was adopted the setting of the three optimization algorithm parameters in Experiment 1, but fixed the factor factorization dimension k=24.After the experiment, the contrast result shows in Chart 3 as below:

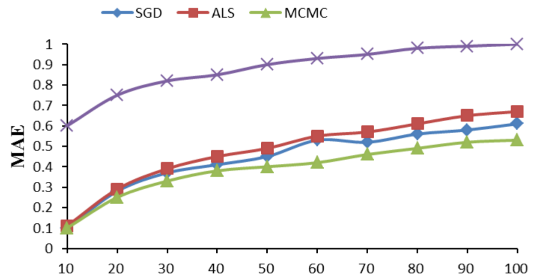

Figure 3. Recommended accuracy under different tourists' sets.

As can be seen from Fig. 3, as the set of tourists increases, the accuracy of recommendation decreases. However, the traditional collaborative filtering algorithms generally have lower recommendation accuracy with the tendency of constantly sparseness. So, its recommendation accuracy is beginning to distort and can hardly be accurately recommended. The performance of the traditional collaborative filtering algorithm in Top_N recommendation ranking accuracy is far lower than that of the FMs model, which shows that the FMs model has better performance on the sparse experimental data set.

Experiment 3: Comparison of recommended labels under different specific gravity factors.

For the sake of comparing the difference of experimental datasets under different specific gravity coefficients, the data of the users about staying at an inn should be taken into account, and the differences between the recommended values of labels under different specific gravity factors are compared. The specific design is as follows:

To suppose $\lambda_1 = 0.6$, $\lambda_2 = 0.4$, you can get the information about the eigenvector values of the same tourist staying in different hotels, showing in Table 4:

According to the formula 24, φ_1, φ_2, φ_3, φ_4, φ_5 and φ_6 respectively, it is represented the hotel star rating (*Hsf*), hotel hardware (*Hhf*), low price ratio (*Lpr*), geographical location (*Gl*), tourist rating (*Te*), and tourist star preference (*Tsp*). Because of the same visitor, the change in the value of φ_6 will not affect the order of recommended tag values, so $\varphi_6 = 0.1$ is fixed here. After that, according to the tourist preference for a hotel in certain aspects or factors, assume that the corresponding weight is 0.7, and the rest is 0.05, then it can get the different recommended tag values after calculating, illustrated in Table 5:

Table 4. Eigenvector values for different hotels.

HotelID	*Hsf*	*Hhf*	*Lpr*	*Gl*	*Te*	*Tsp*
Ht095	0.2	0.1538	0.9557	0.1121	0.2335	0.5551
Ht308	1.0	0.2303	0.9344	0.2256	0.5318	0.5551
Ht311	1.0	0.2036	0.8890	0.2350	0.5011	0.5551
Ht318	0.8	0.2234	0.9221	0.2206	0.5011	0.5551
Ht323	0.8	0.1331	0.9378	0.2438	0.5086	0.5551

Ht326	0.8	0.1221	0.9400	0.2300	0.3700	0.5551
Ht328	0.8	0.1959	0.9776	0.2255	0.4357	0.5551
Ht349	0.6	0.1695	0.9283	0.2364	0.5421	0.5551
Ht360	0.6	0.2911	0.9243	0.2292	0.5889	0.5551

Table 5: Recommended tag values under different specific gravity.

Hotel ID	$\varphi_1=0.7$	$\varphi_2=0.7$	$\varphi_3=0.7$	$\varphi_4=0.7$	$\varphi_5=0.7$
Ht095	0.2683	0.2382	0.7595	0.2111	0.2900
Ht308	**0.8516**	0.3513	0.8090	0.3483	0.5473
Ht311	0.8469	0.3293	0.7748	**0.3497**	0.5227
Ht318	0.7089	0.3341	0.7882	0.3322	0.5146
Ht323	0.7067	0.2732	0.7962	0.3451	0.5173
Ht326	0.6986	0.2580	0.7896	0.3281	0.4191
Ht328	0.7072	0.3146	**0.8227**	0.3338	0.4704
Ht349	0.5693	0.2895	0.7827	0.3330	0.5317
Ht360	0.5772	**0.3764**	0.7880	0.3362	**0.5700**
Ht362	0.5757	0.3665	0.7964	0.3435	0.5395

The highest recommended tag values under different weights can be observed in bold in Table 5. In fact, it is clear from the table that the tourists are biased differently and the ideal recommended reference to meet their individual needs can be obtained through the designed model. This fully reflects that under different weighting factors, different prediction recommendation tags will be obtained, which can fulfill the personalized needs of tourists.

The above experiments show that using the formally defined hotel star factor, hotel hardware facilities, low cost ratio, geographical location, tourist rating and tourist star preference, the recommended label value obtained by the weighted method is defined as the target of the FMs model Vector *Y*. It uses MCMC, SGD and ALS algorithm to accurately recommend tourist hotels according to the needs of tourists, focusing on the precise realization of tourist hotel recommendations. Therefore, it is suggested that tourists should choose the weight coefficient that suits their own needs according to their actual situations, which can be used as the recommended basis of the hotel to obtain the ideal recommendation result, which can effectively improve the quality of tourism.

VI. CONCLUSIONS

This article predicts the tourists' preferences for the tourist hotels from the historical data about staying at an inn, and combines with the many related attributes of the tourist hotels to realize the accurate recommendation of tourist hotels for tourists from different points of view. The recommended accuracy and sorting accuracy are superior to the traditional user-based collaborative filtering algorithm, which finds a new method for the accurate recommendation of intelligent personalized travel hotels. How to better combine the off-season with the peak season of tourist attractions, and the characteristics of the preference of the tourists for the types of touristic attractions and other characteristics for the sake of providing more accurate and diverse tourist hotel recommendations, it will be the next step of the focus in the research.

REFERENCES

Fu-guo, Z., "Survey of Online Social Network Based Personalized Recommendation", *Journal of Chinese Computer Systems*, **35(7)**, 1470-1476 (2014).

Ke, Q., W. Shaofei, M. Wang, and Y. Zou, "Evaluation of Developer Efficiency Based on Improved DEA Model", *Wireless Personal Communications*, **102(4)**, 3843-3849 (2018).

Sun, W., "A Personalized Hotel Recommendation Model Based on User Behavior", *Electronic World*, **24**, 469-469 (2014).

Wu, S., "A Traffic Motion Object Extraction Algorithm". *International Journal of Bifurcation and Chaos*, **25(14)**, Article Number 1540039 (2015).

Wu, S., M. Wang and Y. Zou, "Research on internet information mining based on agent algorithm". *Future Generation Computer Systems*, **86**, 598-602 (2018a).

Wu, S., M. Wang and Y. Zou, "Sewage information monitoring system based on wireless sensor", *Desalination and Water Treatment*, **121**, 73-83 (2018b).

Yu, W. and Li S., "Recommender systems based on multiple social networks correlation". *Future Generation Computer Systems*, **87**, 312-327 (2018).

Received: December 15th 2017
Accepted: June 30th 2018
Recommended by Guest Editor
Juan Luis García Guirao