A Personalized Recommendation Algorithm on Integration of Item Semantic Similarity and Item Rating Similarity

Songjie Gong
Zhejiang Business Technology Institute, Ningbo 315012, China
Email: songjie_gong@163.com

***Abstract*—With the rapid development of the Internet and the wide application of e-commerce, recommender system has become a necessity and collaborative filtering is the most successful technology for building recommendation systems. There are many problems in the recommendation approaches, such as data sparsity problem, the issue of new items and scalability issues. Item-based collaborative filtering algorithms can improve the scalability and the traditional user-based collaborative filtering methods, to avoid the bottlenecks of computing users' correlations by considering the relationships among items. But it still worked poor in solving the issues of sparsity, predictions for new items. In order to effectively solve several problems, this paper presented a recommendation algorithm on integration of item semantic similarity and item rating similarity. The item semantic similarity is calculated combining Earth Mover's Distance and Proportional Transportation Distance, which can utilize the semantic information to measure the similarity between two items based on a solution to the transportation problem from linear optimization1. Then producing recommendation used item-based collaborative filtering integrating the semantic similarity and rating similarity. The presented approach can effectively alleviate the sparsity problem in e-commerce recommender systems.**

***Index Terms*—recommendation algorithm, collaborative filtering, semantic similarity, rating similarity, earth mover's distance, proportional transportation distance**

I. INTRODUCTION

With the development of network technology and e-commerce applications, Internet has emerged the phenomenon of information overload, so people need a personalized recommendation system [1,2]. Personalized recommendation technology is the most important technology services; its objective is to filter out the user is not interested in the item or items of interest to forecast users. Over the past few years, the recommended system is used in many different areas, such as online news filtering, music and movie recommendations, as well as a variety of online product recommendations. To ensure the recommended system to produce accurate real-time recommendations, the researchers propose a variety of recommendation algorithms; the present method is mainly a content-based filtering, Bayesian network technology, association rules technology, knowledge-based recommended methods and clustering technology. Most of the recommendation system there are two drawbacks: first, the data sparseness problem, that is, to establish early in the system, due to system resource has not been enough evaluation of the system difficult to use these evaluations to find similar users; second cold started the problem, that is, if a new project to evaluate its no, then it would not recommend, recommendation system had collapsed.

Collaborative filtering technology is currently the most popular recommendation technology [3,4]. Research in the field of collaborative filtering, there are two main ways: user-based collaborative filtering and item-based collaborative filtering.

User-based collaborative filtering achieved in two steps [5,6]:

(1) first of all, be active users of the k-nearest neighbor set of the user form, this is by calculating the activities of the user model and other models related to each user or similarity to achieve;

(2) for the active user has not evaluated the objective evaluation of the project i generate predictive value.

A major problem with this approach is scalability and sparsity.

The contrary, item-based collaborative filtering method can overcome the user-based extension of the relative problem, it points to achieve the following three steps [5,6]:

(1) system inspection activities the user has evaluated the set of all items N, choose one with the target item ik most similar items (i1, i2, ..., ik) as the impact of projects on the prediction set;

(2) i was calculated by the similarity with the most similar item (Si1, Si2, ..., Sik), influence each similarity. The user sets the value of Sij evaluated by the same time, all users i and j form;

(3) by taking active users of these items to the evaluation of the most similar to the weighted average rating by user i's forecast.

Item-based collaborative filtering algorithms can improve the scalability and the traditional user-based collaborative filtering methods, to avoid the bottlenecks of computing users' correlations by considering the relationships among items. But it still worked poor in solving the issues of sparsity, predictions for new items. In order to effectively solve several problems, in this

doi:10.4304/jcp.6.5.1047-1054

paper, we presented a recommendation algorithm on integration of item semantic similarity and item rating similarity. The item semantic similarity is calculated combining Earth Mover's Distance and Proportional Transportation Distance, which can utilize the semantic information to measure the similarity between two objects based on a solution to the transportation problem from linear optimization1. Then producing recommendation used item-based collaborative filtering integrating the semantic similarity and rating similarity. The presented approach can effectively alleviate the sparsity problem in e-commerce recommender systems.

II. Problem Description

Although the current recommendation system has achieved a wide range of applications, but most do not understand the semantic features of the project. This does not guarantee the accuracy of forecasts. For example, in the strong e-commerce activity in real time, the product may be involved in many industries, such as audio and video products, sports, food, and so on. When the user decides to purchase the product or products to discount the weight, not only on the product or industry representatives interested users, but also the user associated with this commodity is likely to be implicit in the semantic information of interest.

A. Item semantic property

Traditional collaborative filtering is taken into account the user's item level. With the user and item development, the sparsity problem is getting worse. Therefore, we believe that the semantic properties of the item can be used to solve this problem.

The content of many items such as books, movies, or music is difficult to analyze automatically by a computer, but the items may be categorized based on the attributes of the items. For example, in the context of books, every book can be classified according to the semantic attribute of each item. As shown in the figure 1.

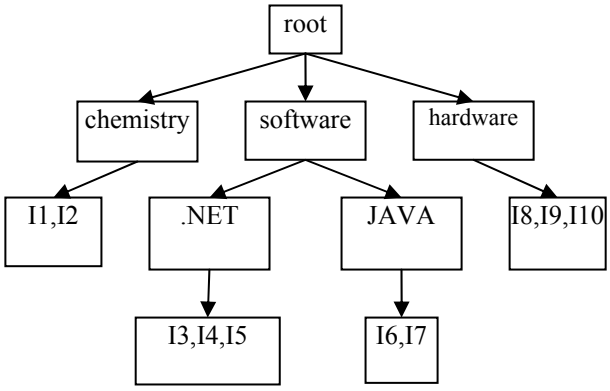

Figure 1. Item hierarchy

B. Problem description

According to user-based collaborative filtering, identifying the target users predicted rating by the neighbors. But it is common to a user a lot of different content, but the history of the item or scores one of the items, and predicted another content item is. In fact the project is not making the prediction history entries.

When a user may be interested in items related to software books, but only rated JAVA software items. Now if we predict hardware item, then content preference of the user for software items will only be used, off course the result is doubtful.

Let's see a detail example as following.

TABLE I
USER-ITEM RATING TABLE

	I1 (software)	I2 (software)	I3 (hardware)	I4 (hardware)
U1	9	9	1	1
U2	8	9	1	2
U3	1	1	9	9
U4	2	1	9	8
U5	1	1	1	?

There are five users and four items in the Table 1 in the user-item rating table. We supposed that, content of item I1 (software) and I2 (software) are software books that have the similar content but are different items. In the same way, content of item I3 (hardware) and I4 (hardware) are hardware books that are different from content of item I1 and I2.

Supposed we will predict rating of user U5 for item I4 and each uses has two neighbor users. If prediction is done according to user-based collaborative filtering algorithm, it is obvious that users U3 and U4 will be neighbors of U5 because of their similar rating behavior to user U5. Then it is easy to get the prediction value. But we find that the reason why users U3 and U4 will be neighbors of active user U5 is mainly that two of them are similar in the all content as software book and hardware book. We used all content of items related to software and hardware to predict, but software book and hardware book are dependent and exclusive, so prediction is not accurate, and collaborative filtering recommender based on prediction is wrong.

III. A Semantic Similary Function Combining EMD and PTD

A. Metric

The difference between the two objects can be the distance function or similarity measure function. If the two objects based on values calculated from the function of the greater difference between them is that the greater the similarity function is the opposite [7,8,9,10]. Similarity between the function and distance function can be interchangeable. Set f is a distance function, range [0, 1] range. It corresponds to the similarity function can be

$$1 - m(\bullet)$$

or

$$1/(1 + m(\bullet))$$

In many applications, m is a need to measure, metric is defined as follows.

Definition 1 Metric

A metric space is a set S, together with a function p: S × S → non-negative real number, so that any of s1, s2, s3 ∈ S, satisfy:

i. if and only if $s1 = s2$, $p(s1, s2) = 0$;

ii. $p(s1, s2) = p(s2, s1)$;

iii. $p(s1, s3) \le p(s1, s2) + p(s2, s3)$;

The function p is called a metric on S.

Generally easier to find a distance function to satisfy the above definition of i and ii features. But the find function to meet the characteristics of iii (triangle inequality) takes some extra effort. Such as object X, Y, Z, and satisfy the triangle inequality of the distance function f, if known, f (X, Z) - f (Y, Z) is greater than a certain number of k, then there is f (X, Y) ≥ f (X, Z) - f (Y, Z)> k, then excluding the X, Y values can be directly inferred distance between f (X, Y) the distance is greater than k.

B. Semantic Structure

Semantic structure is divided into two types: directed rooted tree and directed acyclic graph, referred to as Tree and Diagram. How to select and extract the semantic structure of relationships can be established under specific circumstances [11].

Definition 2 Diagram

Diagram is a directed acyclic graph $G = (V, E, W, r)$;

$V \ne \varnothing$ is the item set;

r is the graph of the root.

$E < \{r\} \bigcup V \times V$ is the edge of the collection;

W is E (G) to the positive real set of functions, called the weighting function;

In diagram G, only r entry is 0;

If the degree is 0, one is based items;

If all the items of V are the base items claimed graph G is trivial.

Definition 3 Tree

Tree is a special diagram. It satisfies: dig up the roots, the entry of the other nodes is 1.

In general, the path diagram in the sequence recorded for the side.

Suppose $G = (V, E, W, r)$ is a graph, with items $x_0, x_1, ..., x_n \in V, n \ge 0$, if any sequence to the edge $P = (e_1, e_2, ..., e_n)$, meet the arbitrary $e_i \in E (1 \le i \le n)$ is x_{i-1} side to x_i, claimed that P is a path from x_0 to x_n.

With the symbol W (P) expresses the sum of the weights of all edges in P. In particular, if there is no path between the two, defined as the empty path, with the $\perp$ that $W(\perp) = 1$.

Definition 4 Semantic path

There is a diagram $G = (V, E, W, r)$. Semantic path of v is a direct path from r to any node $v \in V$. The v has at least one semantic path, their set denoted as $\phi(v)$.

Definition 5 Intersection of semantic path

Given the diagram $G = (V, E, W, r)$, $P = (e_1, e_2, ..., e_m)$ and $Q = (e'_1, e'_2, ..., e'_n)$ are the semantic path of G. May wish to set $n \ge m > 0$, then the intersection of P and Q is a semantic path, denoted $P \bigcap Q$.

Specifically defined: if there is a maximum of $k \ge 1$ meet $e_i = e'_i$, $1 \le i \le k$, $k \le m$ and $e_{k+1} \ne e'_{k+1}$,then $P \bigcap Q = (e_1, e_2, ..., e_k)$; otherwise $P \bigcap Q = \perp$.

C. Distance of items based on semantic path

Definition 6 Distance

Given the diagram $G = (V, E, W, r)$ and notes $u, v \in V \bigcup \{r\}$, the distance of u and v defined as d_Δ.

If $u \ne r$ or $v \ne r$

$$d_\Delta(u, v) = 1 - \max\left\{\frac{2 \times W(P \bigcap Q)}{W(P) + W(Q)} \mid P \in \phi(u), Q \in \phi(v)\right\}$$

If $u = v = r$

$d_\Delta(u, v) = 0$

Given the diagram $G = (V, E, W, r)$ and u , v , $w \in V \bigcup \{r\}$, we can get:

i $0 \le d_\Delta(u, v) \le 1$;

ii if and only if u = v , $d_\Delta(u, v) = 0$;

iii $d_\Delta(u, v) = d_\Delta(v, u)$;

iv if G is a Tree,

$$d_\Delta(u, w) \le d_\Delta(u, v) + d_\Delta(v, w)$$

D. Region

Definition 7 Region

Set X is a set, each map $A: X \to [0,1]$ is called a region of X. $\sum_{x \in X} A(x)$ is the average width of the region

of A, denoted | | A | |. For any $x \in X$, A (x) is the region value of x. When X = $\varnothing$, called empty region. Its average width is 0.

For example, X={<a,1>,<b,0.5>,<c,0.1>} is a region, the region value of a is 1, the region value of b is 0.5, the region value of c is 0.1. Intuitive interpretation, A(x) denotes the size of the possibility of x appeared in the object characteristic

Definition 8 Pre-Region

The Pre-region of region A defined as preA, satisfy:

A=B, if and only if preA=preB, and $x \in \mathrm{dom}\, A, A(x) = B(x)$

Set A is a region, $k \in [0, 1]$, defined $k \times (A)$ as a new region A′, satisfy:

if k=0, A′=NULL,

if $k \neq 0$ $\mathrm{dom}\, A = \mathrm{dom}A'$ and $\forall x(\mathrm{dom}\, A, A(x) = k \times A(x))$

If a region A is satisfying $\forall x \in \mathrm{dom}A, A(x) = 1$, A is a homogeneous region of X.

If a region A is satisfying $\sum_{x \in \mathrm{dom}\, A} A(x) = 1$, A is a regulation region of X.

Definition 9 Match

Given tow non empty region A and B, if a map $M : \mathrm{dom}\, A \times \mathrm{dom}\, B \to [0, 1]$, satisfying:

i. $\forall b \in \mathrm{dom}\, B \sum_{a \in \mathrm{dom}\, A} M(a, b) \leq B(b)$

ii. $\forall a \in \mathrm{dom}\, A \sum_{b \in \mathrm{dom}\, b} M(a, b) \leq A(a)$

iii. $\sum_{a \in \mathrm{dom}\, A} \sum_{b \in \mathrm{dom}\, B} M(a, b) = \min\{ \|A\|, \|B\| \}$

called it as match of region A and B.

Differences in the definition of the sum generated by this match as:

$$\mathrm{DIF}(M) = \sum_{a \in \mathrm{dom}\, A} \sum_{b \in \mathrm{dom}\, B} M(a, b) \times d_{\Delta}(a, b)$$

If there is no other matching M ', made $\mathrm{DIF}(M') < \mathrm{DIF}(M)$, claimed that M is best match of region A and B.

The DIF (M) as the smallest total match, denoted by $\mathrm{minDIF}(A, B)$.

For any two non-null region of A and B, they have the following properties of the smallest total difference.

i. if A=B, then $\mathrm{minDIF}(A, B) = 0$

ii. if $\mathrm{minDIF}(A, B) = 0$ and $\|A\| = \|B\|$ then A=B

iii. $0 \leq \mathrm{minDIF}(A, B) \leq \min(\|A\|, \|B\|)$

iv. $\mathrm{minDIF}(A, B) = \mathrm{minDIF}(B, A)$

Definition 10 Greatest similarity width

Given two regions A and B, the greatest similarity width between them is defined as

if A ≠ NULL and B ≠ NULL

$$A \tilde{\cap} B = \min(\|A\|, \|B\|) - \mathrm{minDIF}(A, B)$$

if A=NULL or B=NULL

$$A \tilde{\cap} B = 0$$

Given tow regions A and B, we can get:

i. $A \tilde{\cap} A = \|A\|$

ii. $0 \leq A \tilde{\cap} B \leq \min(\|A\|, \|B\|)$

iii. $A \tilde{\cap} B = B \tilde{\cap} A$

iv. if $k \in [0, 1]$, then $A \tilde{\cap} (B) \geq A \tilde{\cap} (K \times B) \geq K(A \tilde{\cap} B)$

E. Earth mover's distance

Although the basis functions can be easily adopted to solve the items of the semantic similarity distance between items, but the vector space model in the calculation of similar distance in the document the situation much more complex. Here not only need to calculate the semantic distance between items, but also to find a suitable match items. Similarity algorithm in the past generally been ignored in the matching factors, only the use of the same item in different documents, "one to one" matching. This approach proved intuitive to people close to the effects in poor and vocabulary main reasons synonymous with ambiguity phenomenon. Since different documents may use different words to express the same concept, the "one to one" matching method becomes powerless when dealing with a.

In response to this problem, commonly used in image retrieval EMD algorithm is introduced to the document similarity calculation [12,13,14,15], the proposed use of EMD "many to many" match the characteristics of the document vector integrated semantic matching items, effectively improve the computational accuracy, the following to the a similar document from the definition of EMD:

Definition 11 EMD

Given tow regions A and B, the EMD of A and B is defined as:

if $A \neq \varnothing$ and $B \neq \varnothing$

$$EMD(A, B) = 1 - \frac{A \tilde{\cap} B}{\min\{\| A \|, \| B \|\}}$$

if $A = \varnothing$ and $B = \varnothing$

$EMD(A, B) = 0$

if $A = \varnothing$ or $B = \varnothing$

$EMD(A, B) = 1$

Given three regions A, B and C, we can get property of EMD as follows:

i. $0 \le EMD(A, B) \le 1$

ii. if $\| A \| = \| B \|$, $EMD(A, B) = 0$ then A=B

iii. $EMD(A, B) = EMD(B, A)$

iv. if $\| A \| = \| B \| = \| C \|$, then $EMD(A, C) \le EMD(A, B) + EMD(B, C)$

Definition 12 J-EMD

Given tow regions A and B, the J-EMD of A and B as:

$$J - EMD(A, B) = \begin{cases} 1 - \dfrac{A \tilde{\cap} B}{\| A \| + \| B \| - A \tilde{\cap} B}, A or B \neq \varnothing \\ 0, A = \varnothing and B = \varnothing \end{cases}$$

Given three regions A, B and C, we can get property of J-EMD as follows:

i. $0 \le$J-EMD (A , B) ≤ 1

ii. J- EMD (A , B) = 0 ,if and only if A = B

iii. J- EMD (A , B) = J EMD (B , A)

iv. J -EMD (A , B) + J EMD(B , C) ≥ J- EMD (A , C)

Definition 13 M-EMD

Given tow regions A and B, the M-EMD of A and B as:

$$M - EMD(A, B) = \begin{cases} 1 - \dfrac{A \tilde{\cap} B}{\max(\| A \| + \| B \|)}, A or B \neq \varnothing \\ 0, A = \varnothing and B = \varnothing \end{cases}$$

Given three regions A, B and C, we can get property of M-EMD as follows:

i. $0 \le$M-EMD (A , B) ≤ 1

ii. M-EMD (A , B) = 0 ,if and only if A = B

iii. M-EMD (A , B) = M-EMD (B , A)

iv. M-EMD (A , C) ≤ M-EMD(A , B) + M-EMD (B , C)

F. Proportional transportation distance[16]

Definition 14 PTD

Given tow regions A and B, $\| A \| \le \| B \|$,the PTD of A and B is defined as:

if $A \neq \varnothing$ and $B \neq \varnothing$

$$PTD(A, B) = 1 - \frac{A \tilde{\cap} (\frac{\| A \|}{\| B \|} B)}{\| A \|}$$

if $A = \varnothing$ and $B = \varnothing$

$PTD(A, B) = 0$

if $A = \varnothing$ or $B = \varnothing$

$PTD(A, B) = 1$

Given three regions A, B and C, we can get property of PTD as follows:

i. $0 \le$PTD (A , B) ≤ 1

ii. PTD (A , A) = 0

iii. PTD (A , B) = PTD (B , A)

iv. PTD (A , C) ≤PTD (A ,B) + PTD (B , C)

G. Distance combining EMD and PTD

Recommended application in the local match is required, because the need to predict the future behavior of the user is currently possible behavior [17,18,19]. The percentage distribution of the same characteristics as the same application can be found in the recommended an advantage. For example, if the item with the score as the weight user score may be lower or higher, according to the proportion of matching is a way to solve this problem. But these two features in some applications, the need to avoid. If the session data clustering, in general, local match to should be avoided [20,21,22]. In Web applications recommended weight if the item is check the time, according to the proportion of matches is a bad way. It is necessary to provide a new function to solve these problems.

Definition 15 LLD

Given tow regions A and B, the LLD of A and B is defined as:

$LLD(A,B) = \beta \times PTD(A, B) + (1 - \beta) \times EMD(A, B)$

where, $0 \le \beta \le 1$

if $\beta = 0$,
then $LLD(A,B) = EMD(A, B)$
if $\beta = 1$,
then $LLD(A,B) = PTD(A, B)$

Given three regions A, B and C, we can get property of LLS as follows:

i. $0 \le LLD(A, B) \le 1$
ii. if $\beta > 0$,
then $LLD(A, B) = 0$, if and only if $A = B$
iii. $LLD(A, B) = LLD(B, A)$
iv. $LLD(A, B) + LLD(B, C) \ge LLD(A, C)$

Definition 16 LLS

Given tow regions A and B, the LLS of A and B is defined as:

$$LLS(A,B)=1-LLD(A,B)$$

IV. Combining Item Semantic Similarity and Item Rating Similarity to Produce Recommendations

A. User-item rating table

The task of the traditional collaborative filtering recommendation algorithm concerns the prediction of the target user's rating for the target item that the user has not given the rating, based on the users' ratings on observed items. And the user-item rating database is in the central. Each user is represented by item-rating pairs, and can be summarized in a user-item table, which contains the ratings Rij that have been provided by the ith user for the jth item, the table as following [23,24].

TABLE II
USER-ITEM RATINGS TABLE

Item / User	Item1	Item2	… …	Itemn
User1	R11	R12	… …	R1n
User2	R21	R22	… …	R2n
… …	… …	… …	… …	… …
Userm	Rm1	Rm2	… …	Rmn

Where Rij denotes the score of item j rated by an active user i. If user i has not rated item j, then Rij =0. The symbol m denotes the total number of users, and n denotes the total number of items..

B. Item semantic property structure

The content of many items such as books, videos, or CDs is difficult to analyze automatically by a computer, but the items may be categorized or clustered based on the attributes of the items. For example, in the context of movies, every movie can be classified according to the "genre" attribute of each item. Other item descriptions such as title, category, subject, authors, and published time also reflect the interests of a user when a user reads or downloads items [25,26,27]. Table 3 shows examples of the descriptive information of items.

TABLE III
ITEM-PROPERTY TABLE

Property / Item	P1	P 2	… …	P t
Item1	r11	r12	… …	r1t
Item2	r21	r22	… …	r2t
… …	… …	… …	… …	… …
Itemn	rn1	rn2	… …	rnt

Where, rij denotes the express value of the item to its property. The symbol n denotes the total number of items, and t denotes the total number of item property.

C. Measuring the item rating similarity

There are several similarity algorithms that have been used in the item based collaborative filtering: Pearson correlation, cosine vector similarity, adjusted cosine vector similarity, mean-squared difference and Spearman correlation.

Pearson's correlation, as following formula, measures the linear correlation between two vectors of ratings as the target item t and the remaining item r.

$$sim_1(t,r)=\frac{\sum_{i=1}^{m}(R_{it}-A_t)(R_{ir}-A_r)}{\sqrt{\sum_{i=1}^{m}(R_{it}-A_t)^2\sum_{i=1}^{m}(R_{ir}-A_r)^2}}$$

Where R_{it} is the rating of the target item t by user i, R_{ir} is the rating of the remaining item r by user i, A_t is the average rating of the target item t for all the co-rated users, A_r is the average rating of the remaining item r for all the co-rated users, and m is the number of all rating users to the item t and item r.

D. Measuring the item semantic similarity

We also use the LLS semantic similarity defined in above section.

E. Combining the two similarities

We propose a hybrid method that clusters items by combining the item rating similarity and item semantic similarity. The relative weighting is adopted to adjust the importance of rating similarity and semantic similarity. The integrated measurement of similarity is then derived as following formula.

$$sim(i,j)=h sim_1(i,j)+(1-h)sim_2(i,j)$$

Where, h and 1-h represent the relative importance of the item rating similarity and item semantic similarity, respectively. If h= 0, then the method becomes item semantic-based method. If h=1, then the method becomes traditional item-based collaborative filtering method.

F. Selecting neighbors

Select of the neighbors who will serve as recommenders [28,29]. Two techniques have been employed in recommender systems:

(a) Threshold-based selection, according to which items whose similarity exceeds a certain threshold value are considered as neighbors of the target item.

(b) The top-n technique in which a predefined number of n-best neighbors is selected.

G. Producing Recommendations

Since we have got the membership of item, we can calculate the weighted average of neighbors' ratings, weighted by their similarity to the target item.

The rating of the target user u to the target item t is as following:

$$P_{ut} = \frac{\sum_{i=1}^{c} R_{ui} \times sim(t,i)}{\sum_{i=1}^{c} sim(t,i)}$$

Where Rui is the rating of the target user u to the neighbour item i, sim(t, i) is the similarity of the target item t and the neighbour it user i for all the co-rated items, and m is the number of all rating users to the item t and item r.

V. CONCLUSIONS

With the rapid development of the Internet and the wide application of e-commerce, recommender system has become a necessity and collaborative filtering is the most successful technology for building recommendation systems. There are many problems in the recommendation approaches, such as data sparsity problem, the issue of new items and scalability issues. Item-based collaborative filtering algorithms can improve the scalability and the traditional user-based collaborative filtering methods, to avoid the bottlenecks of computing users' correlations by considering the relationships among items. But it still worked poor in solving the issues of sparsity, predictions for new items.

In order to effectively solve several problems in collaborative filtering, in this paper, we presented a recommendation algorithm on integration of item semantic similarity and item rating similarity. The item semantic similarity is calculated combining Earth Mover's Distance and Proportional Transportation Distance, which can utilize the semantic information to measure the similarity between two objects based on a solution to the transportation problem from linear optimization1. Then producing recommendation used item-based collaborative filtering integrating the semantic similarity and rating similarity. The presented approach can effectively alleviate the sparsity problem in e-commerce recommender systems.

ACKNOWLEDGMENT

A Project Supported by Scientific Research Fund of Zhejiang Provincial Education Department (Grant No. Y201016682).

Programs Supported by Ningbo Natural Science Foundation (Grant No. 2009A610080).

REFERENCES

[1] Breese J, Hecherman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI'98). 1998. 43~52.

[2] Learning Collaborative Information Filters. In Proceedings of ICML '98. pp. 46-53.

[3] Herlocker, J. (2000). Understanding and Improving Automated Collaborative Filtering Systems. Ph.D. Thesis, Computer Science Dept., University of Minnesota.

[4] Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl, J. (2000). Analysis of Recommendation Algorithms for E-Commerce. In Proceedings of the ACM EC'00 Conference. Minneapolis,MN. pp. 158-167

[5] Karypis, G. Evaluation of Item-Based Top-N Recommendation Algorithms. Technical Report CS-TR-00-46, Computer Science Dept., University of Minnesota. 2000

[6] Sarwar B, Karypis G, Konstan J, Riedl J. Item-Based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International World Wide Web Conference. 2001. 285-295.

[7] Christiane Fellbaum. WordNet: An Electronic Lexical Database.Cambridge, MA: The MIT Press, 1998

[8] I M Singer, A Thorpe John. Lecture Notes on Elementary Topology and Geometry. New York: Springer-Verlag, 1987

[9] Olfa Nasraoui, Hichem Frigui, Raghu Krishnapuram, et al.Extracting Web user profiles using relational competitive fuzzy clustering. International Journal on Artificial Intelligence Tools,2000, 9(4): 509～526

[10] Prasanna Ganesan, Hector Garcia-Molina, Jennifer Widom.Exploiting hierarchical domain structure to compute similarity.ACM Trans on Information System, 2003, 21(1): 64～93

[11] Liang Min and Guo Xin-tao,et al..X_dist——a flexiblesemantic distance function. Journal of Computer Research and Development,2004,41(10):1728-1736.

[12] Wan Xiaojun and Peng Yuxin.The earth mover's distance as a semantic measure for document similarity.ACM Fourteenth Conference on Information and Knowledge Management(CIKM),Bremen,2005:301-302.

[13] Rubner Y and Carlo T,et al..The Earth mover's distance asa metric for image retrieval. International Journal of Computer Vision,2000,40(2):99-121.

[14] Rubner Y.Source code for the EMD software. http://robotics.stanford.edu/~rubner/emd/default.htm,Retrieved2007,1

[15] Yossi Rubner.Perceptual metrics for image database navigation.Stanford University,Department of Computer Science,1999.

[16] Panos Giannopoulos, Remco C Veltkarmp. A pseudo-metric for weighted point sets. The 7th European Conf on Computer Vision,Copenhagen, 2002

[17] G Hirst, D St-Onge. Lexical chains as representations of context for the detection and correction of malapropims. In: WordNet: An Electronic Lexical Database. Cambridge, MA: The MIT Press,1998. 305～332

[18] M A Rodriguez, M J Egenhofer. Determining semantic similarity among entity classes from different ontologies. IEEE Trans on Knowledge and Data Engineering, 2003, 15(2): 442～456

[19] Glen Jeh, Jennifer Widom. SimRank: A measure of structural context similarity. The 8th ACM SIGKPP Int'l Conf on Knowledge Discovery and Data Mining, Edmonton, 2002

[20] H Toivonen, M Klemettinen, P Ronkainen, et al. Pruning and grouping discovered association rules. MLnet Familiarization Workshop on Statistics, Machine Learning and Knowledge Discovery in Databases, Heraklion, Crete, Greece, 1995
[21] G K Gupta, A Strehl, J Ghosh. Distance based clustering of association rules. The 9th Artificial Neural Networks in Engineering Conference, St Louis, 1999
[22] A V Goldberg. An efficient implementation of a scaling minimumcost flow algorithm. Journal of Algorithms, 1997, 22(1): 1～29
[23] Che Haoyang, Zhang Jiacai, WWW Collaborative Recommendation Based on Reliability, Journal of Electronics (CHINA), Vol.23 No.2,March 2006,255-258
[24] Duen-Ren Liu, Ya-Yueh Shih, Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences, The Journal of Systems and Software 77 (2005) 181–191.
[25] Chen, Y.-L. et al., A group recommendation system with consideration of interactions among group members, Expert Systems with Applications (2007), doi:10.1016/j.eswa.2007.02.008
[26] Shih, Y.-Y., & Liu, D.-R., Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands, Expert Systems with Applications (2007), doi:10.1016/j.eswa.2007.07.055
[27] Kwok-Wai Cheung, LilyF. Tian, Learning User Similarity and Rating Style for Collaborative Recommendation, Information Retrieval, 2004, 7, 395–410,
[28] Jong-Seok Lee, Chi-Hyuck Jun, Jaewook Lee, Sooyoung Kim, Classification-based collaborative filtering using market basket data, Expert Systems with Applications 29 (2005) 700–704.
[29] Hyung Jun Ahn, A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem, Information Sciences 178 (2008) 37-51.

SongJie Gong was born in Cixi, Zhejiang Province, P.R.China, in July 1, 1979. He received B. Sc degree from Tongji University and M. Sc degree in computer application from Shanghai Jiaotong University, P.R. China in 2003 and 2006 respectively. He is currently a teacher in Zhejiang Business technology Institute, Ningbo, P.R.China.

His research interest includes data mining, information processing and intelligent computing. He has published more than 30 papers in journals and conferences.