

Study on Multi-criteria Personalized Recommendation Algorithm

Luo Ya

Guizhou University of Finance and Economics

Zhao Li

Guizhou University of Finance and Economics

Abstract—Personalized recommendation plays an important role in E-commerce. Single-criteria recommendation algorithm has the features of simplicity and high efficiency but also can easily have the problems of cold startup and sparse data. Multi-criteria recommendation algorithm is to carry out prediction from two prospects: customer and commodity, and fully evaluate and give rates to customer similarity, customer's evaluation for commodity, rank of commodity sales, and commodity similarity, etc. The commodities in a same layer are given an aggregated rate respectively by borrowing the concept of commodity hierarchical tree from aggregation. Eventually the advantages of multi-criteria recommendation algorithm are summarized.

Keywords—E-commerce, Personalized Recommendation, Multi-criteria Algorithm, Aggregate Function

I. INTRODUCTION

E-commerce personalized recommendation is to center on customers, and accurately provide for customers the information and services they are interested in by analyzing customers' personalities, customs and preference so that the needs of each individual customer are satisfied. The essence of the personalization is targeting, that is, using different products or service strategies on different customers. And the essence of recommendation is positivism that is to provide corresponding services based on customers' information and needs. In order to realized the goal, there are many recommendation algorithms proliferated.

II. SINGLE-CRITERIA PERSONALIZED RECOMMENDATION ALGORITHM

Single-criteria recommendation algorithm is to perform calculation based on the information from one aspect and get the recommendation result in the process of data collection and analysis of a recommendation system. The current recommendation algorithms are recommendation based on text content and collaborative filtering recommendation based on customers.

The method based on content is to use information indexing technologies to analyze items' content. Commonly neighbor function and clustering technology are applied to analyze and cluster the text content of the items and the recommendation then is generated based on the item features and customer files. The advantages are simplicity and high efficiency. The modeling and similarity measurement of items can be done offline so the response time for recommendation is short. The shortcomings are that it is difficult to distinguish the quality and style of commodities and it cannot find new commodities and those

interested by customers but only the commodities or services similar to customers' current interests can be found.

Collaborative filtering recommendation determines how to perform analysis and filtering and generate recommendation for the current customer by reference to the selections of other customers who have the similar interests or needs with the current customer. The technology is a way to obtain knowledge and a procedure of extracting formal knowledge from usable knowledge sources. With the facilitation of customer behavior records in the past, it can calculate the similarity of interests and preference between individual customers so as to find out the neighboring customers who have the interests and preference close to a particular person. Then activities can be generated by the suggestions or comments of the customer group composed by those neighboring customers. Though collaborative filtering recommendation can offer recommendation from customers' view, there still exist the problems, e.g. information accuracy, data sparsity, and cold startup, etc.

Single-criteria recommendation algorithm has simple operation data and its operation steps are reduced with high operation efficiency. However, there are still many problems. Therefore, scholars start to consider improving it to be multi-criteria recommendation algorithm.

III. MULTI-CRITERIA PERSONALIZED RECOMMENDATION ALGORITHM

Multi-criteria personalized recommendation algorithm combines the data sources and recommendation ideas from multiple single-criteria recommendation algorithms. It saves the advantages of each single-criteria recommendation and merges the method based on content and the collaborative filtering method in multiple forms so that the quality of recommendation is improved. The mixed recommendation algorithms are commonly used and can be classified into three types: simply combining two methods to generate a recommendation result, combining item information with transaction information to form a single information, and knowledge engineering plus collaborative filtering. Due to the different criteria adopted, the algorithm of three methods can have multiple variations. This paper first establishes multiple criteria for evaluation, then does the rating by multiple criteria, and finally uses the aggregate function to gain the total rate for prediction to generate the recommendation result set.

A. Criteria selection

In E-commerce recommendation, an E-commerce system should take customer features and commodity properties into consideration. Customer features include age,

gender, profession, income, geographic area, and preference, etc. And commodity properties consist of name, classification, brand, price, function, and quality, etc.

Definition 1: Customer feature set U describes all of customer features. Set a specific feature to be $u_i, i=1,2,\dots,n$;

Definition 2: Commodity feature set P describes all of commodity features. Set a specific feature to be $p_i, i=1,2,\dots,n$;

B. Rating by criteria

Customers can set up relation with commodities by certain factors to constitute purchases. E-commerce recommendation can generate a recommendation result set by four relations, including purchase categories of similar customers, customers' evaluation for commodities, best selling commodity list, and commodity similarity. Each relation makes different affects on or has different degrees of association with customer purchase. It is possible to set up corresponding weight sets according to the previous results.

1) CUSTOMER SIMILARITY

Specific groups have similar preference to certain commodities. Not strictly speaking, if two individuals have some or one similar property, they belong to a same category. When categorizing similar customer groups, the association between customers becomes the basis for judgment. The association means the average similarity of customer features.

Calculation of customer similarity:

$$\text{Sim}(U_i, U_j) = \frac{1}{1 + \frac{|u_i - u_j|}{\max(u_i, u_j) - \min(u_i, u_j)}} \quad (1)$$

Among the formula, u_i and u_j are two values of a property U of customers.

2) CUSTOMER EVALUATION FOR COMMODITIES

When a commodity has enough amount of evaluation, the average of the commodity's evaluation can be used to measure the commodity's true quality. And a true customer normally can give his/her correct evaluation for a commodity's essential features, so a customer's evaluation value is close to the commodity's quality, that is, the degree of closeness to the commodity's average evaluation value can be measured by the weight of the commodity's evaluation. The evaluation weight is defined as following,

$$A(u, p) = \frac{\sum_{i \in V_n} (1 - \frac{V_{u,p} - \bar{V}_p}{\max - \min})}{n_u} \quad (2)$$

Among the formula (2), $V_{u,p}$ represents the evaluation value of a commodity p , and \bar{V}_p the average

value of the commodity p 's evaluation. Max and Min are the maximum and minimum value of the commodity p 's evaluation given by customers. n_u is the number of customers u .

3) BEST SELLING COMMODITY LIST

The best selling commodities are those having high sales volume and favored by customers. In the sales volume list of the commodities in one category, those ranked ahead have higher possibility to be selected by customers. So commodities are arranged by the order of sales volume and each has certain weight. Those ranked ahead have larger weight. Set rp to be the weight of best selling, then

$$rp_i = 1 - \frac{n_i}{N} \quad (3)$$

Among the formula (3), n_i represents the sales rank of the commodity i among the commodities in one category. N is the rank of sales volume of all the commodities in the same category.

4) COMMODITY SIMILARITY

The feature similarity of the commodity p_i and p_j is expressed as $\text{simAttri}(p_i, p_j)$, $\text{simAttri}(p_i, p_j) = n$, among which n is the number of the features that the two commodities both have. The larger n is, higher similarity the two commodities have. In actual operation, the different commodities in a commodity directory tree are layered in hierarchy. So the commodity similarity also needs to take commodity hierarchy into consideration. The hierarchical similarity of the commodity p_i and p_j is expressed as

$\text{simClass}(p_i, p_j)$, $\text{simClass}(p_i, p_j) = m$. Therefore, the commodity similarity is,

$$\text{SimP} = \text{simAttri}(p_i, p_j) + \text{simClass}(p_i, p_j) \quad (4)$$

C. Multi-criteria recommendation

1) COMMODITY RECOMMENDATION OF SIMILAR CUSTOMERS

The intersection set of purchase commodities of similar customers is used to determine the recommended commodities for customers in the same category. The algorithm is,

Step 1: Compute customer similarity;

Step 2: Get the commodity set of customers who purchased;

Step 3: Compare new customer's search words for purchase;

Step 4: Recommend commodities.

Set the recommended set of the method to be p_{su_i} .

2) AGGREGATE FUNCTION RECOMMENDATION

The aggregate function essentially is to integrate the recommendation results by all criteria and then make a final recommendation that is the multiplication of the recommendation results by all criteria and the weights. That is, the method combines commodity recommendation by similar customers, recommendation by customer evaluation

as well as similar commodities, and best selling rank of commodities, to make a recommendation of commodities. Set the final recommendation result to be $O(p)$, then:

$$O(p) = w_i * P + A(u, p) * P + \text{simAttri}(P_i, P_j) * P + r_{p_i} * P \quad (5)$$

The algorithm of the recommendation is as following:

Input: Targeted customer u , commodity information table, customer table, sales records, and preliminarily predicted rating list for the commodities being rated by customers.

Output: Recommended commodity list

Step1: Search all customers except the targeted customer from customer data table and set as

$$u = \{u_1, u_2, u_3, \dots, u_n\};$$

Step 2: while the set U is not empty, do

Compute the commodity set evaluated by u and u_i and write as P_u and P_{u_i} respectively, then get the union set $P_{uu_i} = P_u \cup P_{u_i}$ of the commodities evaluated by u and u_i .

Find out the commodity set $N_u = P_{uu_i} - P_u$ that is not evaluated by u .

while N_u is not empty, do

As for any commodity m in N_u , the preliminary prediction rating for the commodity m can be found in the preliminary prediction rating list for the commodities not given any rating by customers.

Compute the similarity between u and u_i , $\text{SimP}(u, u_i) = \text{simAttri}(u, u_i) + \text{simClass}(u, u_i)$.

endwhile;

Step 3: select the n customers with the largest sim values of the targeted customer u to form a similar customer group for the customer u .

$$u = \{u_1, u_2, u_3, \dots, u_n\}$$

Step 4: Search the commodity set that the targeted customer u does not evaluate from the

commodity information table and customers rating history.

$$N_u = \{P_1, P_2, P_3, \dots, P_n\}$$

Step 5: while N_u is never empty, do

Compute the average rate of all the commodities in the category that the commodity belongs to given by the targeted customer.

endwhile;

Step 6: Use Formula (5) to compute the total rate.

Step 7: Select the n commodities with the largest rates and recommend them to the customer u

IV. CONCLUSION

Multi-criteria recommendation algorithm has no problem of cold startup. When there is no customer record in the system, it is able to recommend by commodities' similarity. The method perfectly solves the problem of data sparsity. Customer evaluation may be extreme or not given, causing the data to be sparse. In such circumstances, the multi-criteria recommendation algorithm can compute by the average value of customer evaluation so as to make the customer evaluation convergent.

REFERENCES

- [1] William Cheetham, Global Rate Selector: Areconunender System for Supporting the Sale of Plastic Rssin [C], ICCBR2003, P:96~106, 2003
- [2] Kyung-Yong Jung, Jun-yeog Choi, Kee-Wook Rim, and Jung-Hyun Lee, Development of Design Recommender System Using Collaborative Filtering[C], ICADL 2003.p:100-110
- [3] Deng Ailin, Research on Recommending System of Primary Technologies in E-commerce [D], Shanghai: Fudan University, 2003
- [4] Yu Li, Liu Lu, Luo Zhanghua, Comparison and Analysis on E-Commerce Recommendation Method in China[J], System Engineering – Theory & Practice, 2008.P:96 — 101
- [5] Shi Zhongzhi, Knowledge Discovery [M], Tsinghua University Press, 2002
- [6] Yu Li, E-commerce Personalization – Theory, Methods & Applications [M], Tsinghua University Press, 2007
- [7] Hu Bin, Research and Design of Collaborative Filtering Recommendation System Based on Customer and Resource Weight, Master Thesis of Wuhan University of Technology, 2009
- [8] Zhang Yu, E-commerce Recommendation System Based on Product Feature and Hierarchy, Master Thesis of Central China Normal University, 2008.