Journal of Discrete Mathematical Sciences and Cryptography

ISSN: 0972-0529 (Print) 2169-0065 (Online) Journal homepage: https://www.tandfonline.com/loi/tdmc20

A hybrid recommendation algorithm–based intelligent business recommendation system

Fan Yang

To cite this article: Fan Yang (2018) A hybrid recommendation algorithm–based intelligent business recommendation system, Journal of Discrete Mathematical Sciences and Cryptography, 21:6, 1317-1322, DOI: 10.1080/09720529.2018.1526408

To link to this article: https://doi.org/10.1080/09720529.2018.1526408

Published online: 30 Oct 2018.

Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972–0529 (Print), ISSN 2169–0065 (Online)
Vol. 21 (2018), No. 6, pp. 1317–1322
DOI : 10.1080/09720529.2018.1526408

Taylor & Francis
Taylor & Francis Group

A hybrid recommendation algorithm–based intelligent business recommendation system

Fan Yang
New Media College
Communication University of Zhejiang
Hangzhou
Zhejiang
China

Abstract

In order to solve the current issues of e–commerce recommendation systems: low accuracy, inflexible, lack personalized, etc. a solution based on hybrid recommendation algorithm is proposed, aiming at building a personalized recommendation system for e–commerce. To make up for the lack of a single recommendation algorithm, we implement the hybrid recommendation algorithm including three algorithms: content based recommendation algorithm, item based collaborative filtering recommendation algorithm and demography based recommendation algorithm. To expand recommendation dimensions, we adopt the algorithms in classification and clustering to mine the historical data of items and users. Then we do the performance evaluation and the result shows that the improved scheme has a better effect of recommendation for E–commerce.

Keywords: *Recommendation system, Hybrid recommendation, Intelligent recommendation, Collaborative filtering, Data mining*

1. Introduction

The rapid development of e–commerce makes life easier for consumers. However, users feel hard to search valuable information from the vast. As the e–commerce recommendation system [1], the core task is providing personalized recommendation by exploring the connection between the user and the goods and eventually achieving a win–win situation in users and merchants [2]. Nevertheless, e–commerce recommendation systems in domestic have the following drawbacks [3, 4]: 1) recommendation algorithm is too single to meet the individual demands of users;2) the precision of recommendation is too low since *t* he dimensions of user and product can't be mined meticulously; 3) the fixed model and parameters makes recommendation system inflexibility.

In this paper, we design an intelligent business recommendation system. As to enhance the precision of recommendation and meet the individual needs, we mix the recommendation algorithms including content based recommendation algorithm, item based collaborative filtering recommendation algorithm and demography based

E–mail: **yangfan@zjicm.edu.cn**

recommendation algorithm, and use the algorithms in classification and clustering for the historical data of items and users.

2. The design of intelligent business recommendation system

Recommendation module pops data from order library, user label library and commodity tag library, and then calculates by the recommendation algorithm. In the process of recommendation module calculating, it needs to get data updates in real time from commodity pages so that the information provided to downstage page consistent with the actual information. Recommendation results will return to downstage page by the invocation. In this recommendation module, wemix the multiple recommendation results which come from multiple recommendation algorithms, and then filter and rank them as initial results. This design can effectively control the separate recommendation algorithm not to affect the results. It also can adjust the weights of various recommendation algorithms flexibly.

By the recommendation module, adopting content based recommendation algorithm needs to know commodity keywords as well as adopting demography based recommendation algorithm needs to know the basic information of users, but these information always are incomplete. And the ranking procedure of recommendation module also needs preference of users for sorting. In order to support the achievement of the recommendation algorithms and increase the recommendation dimensions, we design a recommendation dimensions updated module in this recommendation system. It is responsible for mining recommendation dimensions from the historical data of users and commodities, and updates the result in the user label library and commodity tag library for recommendation module calling.

3. The implementation of recommendation module

In this recommendation module, we mix three recommendation results which come from three recommendation algorithms, namely content based recommendation algorithm, item based collaborative filtering recommendation algorithm and demography based recommendation algorithm, and then filter and sort them as initial results. We also optimize the three kinds of recommendation algorithm to make them more suitable for e–commerce system.

3.1 *Item based collaborative filtering recommendation algorithm*

Main steps of this algorithm: first, calculate the similarity between items;second, generate recommendation list by the similarity between the items and the historical data of user preferences.

There are some hyperactive users in the e–commerce system, and they affect the calculation of similarity between the items. We optimize the calculation of similarity by using the equation (1) to reduce the impact of hyperactive users on the system accuracy:

$$W_{ij} = \frac{\sum_{u \in N(i) \cap N(j)} \frac{1}{\log 1 + |N(u)|}}{\sqrt{N(i)N(j)}} \quad (1)$$

W_{ij} stands for the similarity between item i and j, $N(i)$ stands for users who prefer item i, $N(j)$ stands for users who prefer item j.

After getting the similarity, we make $p(u, j)$ standing for how the user u like the item j, $W(j, k)$ standing for k items set which is the most similar with item j. $N(u)$ says the items set which is user u is fond, and S_{ij} says the similarity between item i and item j. R_{ui} indicates that user u interested in items i, and R_{ui} can equal 1 if you don't use the rating data. Using the following equation, and then sorting the values, we can conclude the recommendation result:

$$P(u, j) = \sum_{i \in W(j,k) \cap N(u)} S_{ij} * R_{ui} \quad (2)$$

3.2 *Content based recommendation algorithm*

This algorithm can be divided into three steps: first, get the keywords for each item; second, calculate the similarity between keywords; in the end, get how users like the items.

The content of an item can be represented as the following form;we make the k_i standing for the i^{th} keyword of item o_i, w_{ij} standing for the weight of k_i on o_j, and then the content vectoring model of item o_j can be defined as the following formula:

$$\text{content}(o_j) = \{w_{1j}, w_{2j}, w_{3j}, \ldots\} \quad (3)$$

Users can be fond of multiple objects, so the user's preference vector is the vector model average of all user preferences. Take $N(u)$ to be a set of items which user u like, then the user preference vector is defined as following equation:

$$\text{user profile}(u) = \frac{1}{|N(u)|} \sum_{o \in N(u)} \text{content}(o) \quad (4)$$

Given keyword vector of the item, similarity between items can be measured according to the similarity between keywords, and then we can get the similarity between items by cosine calculation. Set W_{ij} to be the similarity between two contents, d_i and d_j stands for the keyword vectors of item i and item j, using the following equation:

$$W_{ij} = \frac{d_i * d_j}{\sqrt{|| d_i || || d_j ||}} \quad (5)$$

After get the similarity, according to the equation (3), recommendation values are obtained.

3.3. *Demography based recommendation algorithm*

Implementation of the proposed algorithm has the following several steps: firstly, get the user demographic information; secondly, classify the users according to demographic information; finally, recommend the favorite goods of other users in this category for the user.

The core of demography based recommendation algorithm is calculating each characteristic user's favorite items. Take $N(i)$ to be a set of users who prefer item i and $U(i)$ to be a set of users who has the character is t icf. $P(f, i)$ is defined as the popularity of item i among the set of users who has the characteristic f, using the following equation:

$$P(f,i) = \left|N(i) \cap U(f)\right| \tag{6}$$

Using the equation(6) leads that a hot commodity will get high ranking in each user category, so in order to launch more different types of goods, the equation is modified to:

$$P(f,i) = \frac{\left|N(i) \cap U(f)\right|}{\left|N(i)\right| + \alpha} \tag{7}$$

In this equation, αmainly solves the problem of data sparseness. If a product is bought by only one user, then $P(f, i)$ is 1, however, plus α can prevent this kind of situation.

4. The implementation of recommendation dimensions updated module

Recommendation dimensions updated module contains two parts: commodity dimension update and user dimension update. It is responsible for mining recommendation dimensions from the historical data of users and commodities, and provides update information for recommendation module.

4.1 *User dimension updated module*

User dimension is divided into identity dimension, demographic dimension and user preference dimension.

For the extraction of users' gender, we adopt support vector regression algorithm [9].The basic thought of support vector regression is making the observations of a sample space mapped to high–dimensional feature space through a nonlinear mapping, and taking linear regression in this feature space. Finally it is concluded that the SVR regression estimate function is:

$$f(x) = \sum_{x_i \in SV} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{8}$$

Because this algorithm needs to set parameter, in order to get the optimal parameters, we should draw a small amount of data for parameter tuning at first time. Model by SVR function after getting the optimal parameters, and then find the predictive value of user's name and gender by this model. Add those two predictive values, if greater than 0 the result is male users, if less than 0 the result is female users.

Support vector regression algorithm is equal for the calculation of each dimension, but not all purchase categories can be used as the key to distinguish gender, we should

find some purchase categories which can clearly distinguish gender and increase the weight of these categories, and then update the model to get a better result. Therefore, we adopt ID3 algorithm of information gain [10] to select categories which own apparent gender variations, and increase the user's purchase weights of these categories by adding the number of purchase.

4.2 *Commodity dimension updated module*

Commodity dimension is divided into basic attribute, commodities general dimension and the search term.

For completing the tags of commodity, at first we adopted Bayes classification, but Bayes classification depends so much on various probabilities that the effect of classification is not ideal. After that, we decide to use the Fisher classifier [11, 12] and obtain abetter effect.

For dividing the grades of commodity, we take the K–means clustering algorithm [11, 12] to calculate which kinds of grade a commodity belong to by the price of it. But K – means algorithm has a few drawbacks, like K value selected with difficulty, the clustering results affected by the initial clustering center and noise data. Aiming at these shortcomings, we make the following adjustments:

1) Take the price data of commodity, and get rid of the noise of this data. We adopt the way of Z–score in the statistics to remove noise data.
2) Data normalization.
3) Set the initial clustering center. In the design of K – means algorithm, set the center of the initial data to five points of same span.
4) Execute K–means algorithm, until the center no longer changes get the class value.

5. Performance Evaluation

There are some key metrics measuring the performance of intelligent business recommendation system: recall ratio, precision, coverage ratio and new item ratio [13].

In this paper, the off–line experimental data are taken from users' shopping data of dangdang.com during May to December 2012. Compared hybrid recommendation algorithm with three single recommendation algorithms namely content based recommendation algorithm, item based collaborative filtering recommendation algorithm and demography based recommendation algorithm, it is concluded that experiment data of recall ratio, precision, coverage ratio and new item ratio.

Known by above data, the precision of item based collaborative filtering recommendation algorithm is high, but the new items ratio is low; The precision of content based recommendation algorithm is less than collaborative filtering algorithm, however, the recall ratio and coverage ratio of this algorithm are highest and it can recommend a lot of new products; The precision of demography based recommendation algorithm is lowest, while the coverage ratio and new product ratio are really high; And the proposed hybrid recommendation algorithm is overall outstanding in precision, recall ratio, coverage ratio and new product ratio.

6. Conclusion

We propose the design and implementation of an intelligent business recommendation system based on hybrid recommendation algorithm, and it provides reference for e–commerce recommendation system. We design overall architecture of this recommendation system with hybrid recommendation algorithm as the core. This system includes the recommendation module and recommendation dimensions update module. In the recommendation module, we introduce the optimization and implementation of content based recommendation algorithm, item based collaborative filtering recommendation algorithm and demography based recommendation algorithm. In the recommendation dimensions update module, we introduce mining recommendation dimensions from user data and commodity data by the algorithms in classification and clustering. In the end, this intelligent recommendation system is proved more efficient than other single recommendation algorithm by key metrics. This system can validly solve the problems existing in the domestic e–commerce recommendation system, and support e–commerce enterprisein the era of personalized consumption.

7. Acknowledgement

This work was supported by Zhejiang province public welfare Technology Application Research Project, granted by Science Technology Department of Zhejiang Province, P.R.China(Grant No.2016C31086)

References

[1] Chang C H. A study on the behavioral intention of implementing RFID System – Taiwan logistics as an example[J]. *Journal of Information & Optimization Sciences*, 2010, 31(2):413–427.

[2] Ting C W, Chen M S, Lee C L. E–satisfaction and post–purchase behaviour of online travel product shopping[J]. *Journal of Statistics & Management Systems*, 2013, 16(2–3):223–240.

[3] Wang H F,Wu C T.A strategy–oriented operation module for recommender systems in E–commerce [J]. Computers & Operations Research,2012,39 (8):1837–1849.

[4] Wei K,Huang J, Fu S.A survey of e–commerce recommender systems [C]//International Conference on Service Systems and Service Management. IEEE,2007:1–5.

[5] Sarwar B, Karypis G, Konstan J. Item–based collaborative filtering recommendation algorithms. 2001

Received April, 2018

Revised June, 2018