CHEMICAL ENGINEERING TRANSACTIONS
VOL. 51, 2016
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Tichun Wang, Hongyang Zhang, Lei Tian

ISBN 978-88-95608-43-3; **ISSN** 2283-9216
DOI: 10.3303/CET1651128

Enterprise Online Product Recommendation Service Model based on Big Data Environment

Weiwei Liu

Northeast Petroleum University, Qinhuangdao 066004,china
huntercet@qq.com

With the development and application of e-commerce, the research on enterprise online product recommendation service model under big data background has become a frontier issue. Collaborative filtering algorithm is improved based on domain ontology, which calculates semantic similarity of domain ontology from two angles of hierarchical similarity and attribute similarity. It combines with the traditional grading similarity to dig out semantic relationship between products, and then it draws abstract semantic information. The experiment results show that it can significantly improve the recommendation speed. Besides, recommendation efficiency is also relatively stable. In dealing with large data, computational efficiency is better than the traditional collaborative filtering algorithm, recommendation algorithm based on association rules and recommendation algorithm based on content.

1. Introduction

With the development and application of e-commerce, the research on personalized information recommendation service model under big data background has become a frontier issue (Junyean Moon, et. al, 2008). E-commerce websites not only provide products and services, but also make users more difficult to find the product information, which meet their needs in mass of information quickly and accurately. Personalized information recommendation based on big-data could recommend products or services to users according to their preference actively in real time. On the one hand, it can better meet the users' individual needs. On the other hand, it could help electric website to establish a stable user groups, improve service quality, and thereby enhance market competitiveness (Lee et al., 2012). As the time for big data processing is coming, several problems that traditional recommendation system faced such as cold start up, accuracy, and scalability are worsen. In the meantime, the real-time problem as a new bottleneck to the recommendation systems oriented huge amounts of data under this severe environment arises. How to provide a better user experience under the big data environment continues to drive the development of technique. With the combination of the distributed system and the grid computing, Cloud computing has developed. The powerful capabilities of preserving and processing big data meet the demand of the recommender system in the big data era. So, how to optimize and parallelize the mature technique in the traditional recommendation system becomes a new area of research (Hu, et al., 2014).

At present, collaborative filtering technology is widely used in personalized recommendation technology (Zhu et al., 2014). After comparing the advantages and disadvantages of the existing recommendation algorithms (Zhu et al., 2014; Long et al., 2012), according to the characteristics of the product recommendation platform and system goal, we choose the collaborative filtering recommendation algorithm. Collaborative filtering algorithm is not perfect, however, it still needs further improvement in the late, in order to meet the accuracy and speed requirements of the recommendation system (Kuang, 2012; Papadimitriou and Disco, 2008).

The study object of this article is enterprise online product recommendation service model of e-commerce based on big-data. It introduces related theoretical basis, including concepts, features and primary technology. It constructs a model for e-commerce enterprise online product recommendation influence factors and then test and revises it according to the data of the questionnaire survey, which is designed based on factors that influence consumers to buy personalized commodity. At the same time, it constructs an enterprise online

Please cite this article as: Liu W.W., 2016, Enterprise online product recommendation service model based on big data environment, Chemical Engineering Transactions, 51, 763-768 DOI:10.3303/CET1651128

product recommendation service model and proposes a model for e-commerce based on achievements of scholars both abroad and at home and conclusion by questionnaire survey. Lastly, it carried on case studies for enterprise online product recommendation service model, through specific analysis of shopping website, to illustrate the feasibility and effectiveness of this model.

This paper proposes a specific enterprise online product recommendation service model based on big-data, accelerating personalized and intelligent development of e-commerce information services. It has brought convenience and personalized service to users and huge economic benefits for e-commerce enterprise and has great significance to personalized information service for e-commerce. In this paper, customer evaluation algorithm based on data mining mainly includes the following several aspects. In the next section, principle of collaborative filtering recommendation is investigated. In Section 3, improved collaborative filtering recommendation based on domain ontology is proposed. In Section 4, in order to test the performance of proposed recommendation algorithm, it is used to evaluate customer credit of some rural credit cooperatives in China. Finally, some conclusions are given.

2. Collaborative filtering recommendation

This algorithm produces recommendation results for the client demanding products. The main idea is to analyze the user data to the evaluation of products. Through the similarity, it finds the nearest neighbor client, and it recommends product for target users based on the nearest neighbor users. Collaborative filtering recommendation algorithm is mainly divided into three steps, the similarity between the user and choosing the nearest neighbor users and recommendation based on predicting scores.

Similarity calculation should calculate personal information of users, evaluation data of products and browsing data. The score can use user-product matrix. If evaluation vector of user a and user b are $\vec{a}$ and $\vec{b}$ respectively, the similarity between user a and user b is

$$sim(a,b) = \frac{\sum_j^n R_{a,j} R_{b,j}}{\sqrt{\sum_j^n R_{a,j}^2} \sqrt{\sum_j^n R_{b,j}^2}} .$$

$R_{a,j}$ and $R_{b,j}$ represent the score of user a and user b to product j respectively. After the completion of the similarity calculation between the users, the similarity is represented by vector length. The shorter the length of the vector, the higher the similarity. In the selection of the nearest neighbor, you have three ways. The similarity threshold can be set, and users satisfying the similarity threshold are the nearest neighbors. We can also set the number of the nearest neighbors. You can also choose some nearest neighbors meeting similarity threshold. At last, the recommendation results are generated. Suppose there are a number of users, K number of nearest neighbors meeting the threshold value. $p_{a,t}$ represents score of user a to product t.

$$P_{a,t} = \overline{R_a} + X \sum_{u=1}^{K} sim(a,u)(R_{u,t} - \overline{R_u}) .$$

sim(a, u) represents similarity between user a and its nearest user u. $\overline{R_u}$ represents average score of neighbor user u to the product and $\overline{R_a}$ represents average score of a to the product. $R_{u,t}$ represents evaluation score of user u to product t.

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, creation, search, sharing, storage, transfer, visualization, and information privacy. The term often refers simply to the use of predictive analytics or other certain advanced methods to extract value from data, and seldom to a particular size of data set.

Analysis of data sets can find new correlations, to "spot business trends, prevent diseases, and combat crime and so on." Scientists, practitioners of media and advertising and governments alike regularly meet difficulties with large data sets in areas including Internet search, finance and business informatics. Scientists encounter limitations in e-Science work, including meteorology, genomics, complex physics simulations, and biological and environmental research. Data sets grow in size in part because they are increasingly being gathered by cheap and numerous information-sensing mobile devices, aerial (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks. The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s; as of 2012, every day 2.5 Exabyte's (2.5×1018) of data were created; The challenge for large enterprises is determining who should own big data initiatives that straddle the entire organization.

Relational database management systems and desktop statistics and visualization packages often have difficulty handling big data. The work instead requires "massively parallel software running on tens, hundreds,

or even thousands of servers". What is considered "big data" varies depending on the capabilities of the users and their tools, and expanding capabilities make Big Data a moving target. Thus, what is considered to be "Big" in one year will become ordinary in later years. "For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options. For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration."

This location data retrieved from smart phones can play a vital role in determining the user trend. As mobile phone is connected with the base station where each base station represents the cell where user is residing in at a particular time it record the spatiotemporal trend of user automatically even without disturbing the user routine. So we can apply the data mining techniques on such kind of data to extract meaningful information. As this location information can provide the list of significant locations for user during mobility this can be used for Location based services (LBS).

The location information and mobility path can be used for the potential applications which include mobile advertisement, city wide route sensing, region pollution, traffic safety management, social networking, potential warning system, soured analysis, route tracking expand and communication. In these applications the low level mobility data is unity recommendations interoperated into high level information in term of stay locations, patterns and finally user profiling.

In our work we are focused on the precise extraction of mobility profile against user mobility so that it can be rendered for any location based service. As described before this mobility profiling is all location based where location is logged in by the user using different methods i.e. Indoor and Outdoor.

There are many ways to record the user mobility which can be Wi-Fi, Bluetooth, Infrared, GPS and GSM depending on the situation and type of intended application. Most important work regarding the location extraction based on algorithms is done in their work where they formally defined the term mobility mining to extract patterns through profiling. While the current mobility trends are studied in detail by their work, the mobility is defined as key prediction indicator of human life.

So the tracking of true location is basis of every mobility based application. As there are two kind of technical solutions available for the location recording indoor and outdoor. In case of indoor like Bluetooth, RFID or infrared, these short range and cannot compete the outdoor like GSM and GPS which are categorized broadly in their work. On the other hand Wifi is another solution to location tracking as well were geo location of the user is determined by the terminal it is connected with. As per the feasibility and their wide usage outdoor technologies are widely used for location tracking which includes GPS, Assisted faux GPS and GSM. Where GPS is coordinated based which provides the exact location of the user in term of longitude and latitude. But GPS needs long start-up time on device, high consumption of energy which is discouraging for the user. And most importantly there are many applications where exact location of user does not matter and application can use the relative position of the user for prediction of trends where GSM can serve the purpose well enough.

3. Improved collaborative filtering recommendation based on domain ontology

The constructed domain is product domain ontology. After ontology knowledge base construction has been completed, the corresponding product characteristics tree of each product is generated. According to feature attributes of the products, we find the location in the knowledge base. After position is determined, according to the position information of product in the domain ontology knowledge base, similarity of product is calculated. On the one hand, it can solve the drawback of ignoring the semantic relationship between keywords in traditional collaborative filtering algorithm to enhance the accuracy of recommendation. On the other hand, the establishment of a comprehensive domain knowledge base, positions of each attribute and attribute value are fixed. For the products, it only needs to store location information of the corresponding position and does not need to inquiry, deal with and storage keywords information in the form of text. It can improve data processing speed. Then Hierarchical similarity calculation and attribute similarity calculation is investigated.

It constructs a model for e-commerce enterprise online product recommendation influence factors and then test and revise it according to the data of the questionnaire survey, which is designed based on factors that influence consumers to buy personalized commodity. At the same time, it constructs an enterprise online product recommendation service model and proposes a model for e-commerce based on achievements of scholars both abroad and at home and conclusion by questionnaire survey. Lastly, it carried on case studies for enterprise online product recommendation service model, through specific analysis of shopping website, to illustrate the feasibility and effectiveness of this model. How to provide a better user experience under the big data environment continues to drive the development of technique. With the combination of the distributed system and the grid computing, Cloud computing has developed.

If A and B belongs to the same branch class, semantic distance between A and B is

$d(A,B)=dep(B)-dep(A)$.

dep(x) represents the depth of class x in the hierarchical structure. If A and B belongs to the heterogeneous class, semantic distance between A and B is
d(A,B)=dep(A,R)+dep(B,R).
The semantic similarity between A and B is
Csim(A,B)=1/(d(A,B)^2+1).
Hierarchical similarity between instance I_1 and I_2 is

$$Isim(I_1, I_2) = \begin{cases} 1, & if\ I_1 = I_2 \\ \dfrac{Csim(C(I_1), C(I_2))}{2}, & otherwise \end{cases}.$$

$C(I_1)$represents class that instance I_1 belongs to. C (I_2) represents class that instance I_2 belongs to. Suppose ontology class C_1 has instance I_1, value of attribute M_1 is m_1, and value of attribute M_2 is m_2, which can be represented as I_1= C_1[M], M=(m_1,m_2,L,m_n)
Ontology class C_2 has instance I_2, value of attribute N_1 is n_1, and value of attribute N_2 is n_2, which can be represented as I_2= C_2[N],N=(n_1, n_2,L, n_n). Attribute similarity between I_1 and I_2 is

$$Psim(I_1, I_2) = \frac{\sum_i \beta_i \cdot com(p_i, q_i)}{\sum \beta_i}.$$

$$com(m_i, n_i) = \begin{cases} 1, & m_i = n_i \\ 0, & otherwise \end{cases}$$, β_i represents weight of some attribute in the class.

$$sim(a,b) = \frac{m \cdot Isim(a,b) + n \cdot Psim(a,b)}{m+n}$$

m and n are constants, Csim(a,b) represents hierarchical similarity between a and b, and Psim(a,b) represents attribute similarity between a and b.

4. Experiment and analysis

To further enhance performance of recommendation algorithm, relying only on a computer is difficult to implement, so parallel processing will effectively improve computing speed of the algorithm. More current application to solve big data includes MapReduce programming model, Hadoop distributed framework and Hbase distributed database (He et al., 2008; Xia et al., 2011; Shekhar et al., 2012; Yang et al., 2011). Here we select MapReduce which is relatively simple and convenient to further improve the algorithm to increase the speed of recommendation. It includes three processes of data division, Map stage and Reduce stage. We mainly carry out segmentation in accordance with the user. Similarity calculation process of a particular user and other users, prediction score and recommendation process are encapsulated in the process of the map.
One hundred of users evaluate 1700 kind of products, and scoring value is an integer from 1 to 10. Two thousand of data are selected. User interest is proportional to scoring value. Accuracy comparison of different algorithms is shown in figure 1. From the perspective of recommendation accuracy, mean absolute error MAE is used to measure the accuracy of the algorithm. MAE is average of deviation of absolute value between actual value and predictive value of all score. The higher the prediction precision, the smaller the MAE value. Product recommendation algorithm accuracy is measured based on MAE. Collaborative filtering recommendation based on domain ontology, traditional collaborative filtering recommendation, recommendation based on the content, and recommendation based on association rules are tested. The experimental results show that the collaborative filtering recommendation algorithm based on domain ontology has higher accuracy than the traditional collaborative filtering recommendation, recommendation based on association rules, and recommendation based on content. Speed comparison of different algorithms is shown in figure 2. It can be concluded that with increasing of number of data, processing speed is much faster than other traditional algorithms. The proposed algorithm has important practical significance for the implementation of enterprise online product recommendation service model.

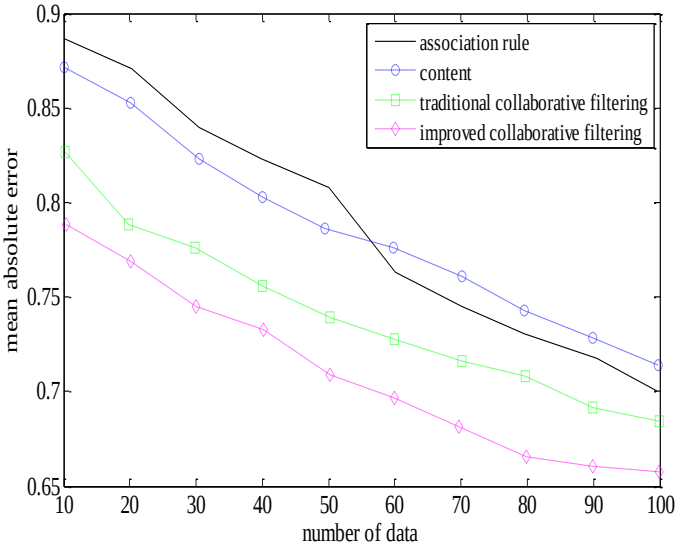

Figure 1: Accuracy comparison of different algorithms

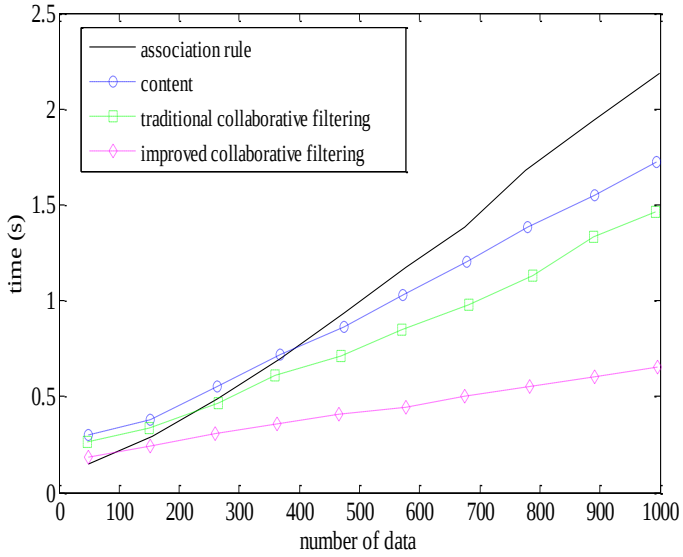

Figure 2: Speed comparison of different algorithms

5. Conclusion

Combined with the characteristics of big data, on the basis of analysis of the current research status of the enterprise online product recommendation system, collaborative filtering recommendation algorithm is selected for its lower data requirements and more successful application. The collaborative filtering recommendation algorithm is difficult in handling the natural language expressive and potential user needs under the environment of big data, which would affect the recommendation accuracy. In order to solve this defect, semantic similarity is introduced into the traditional collaborative filtering algorithm. Domain ontology is introduced to improve performance of enterprise online product recommendation algorithm. It improves the accuracy of recommendation, taps the dynamic and potential needs of users. Besides, it transforms the user needs expressed by the text into location information of the ontology knowledge library; therefore it improves the speed of recommendation. In order to test the effectiveness of the improved algorithm, experiments are done. Testing results show that the improved algorithm can improve recommendation efficiency and recommendation quality to some extent.

Acknowledgment

This work is supported by Qinhuangdao Science and Technology Bureau “The development of medium-sized and small enterprises in Qinhuangdao in the background of big data”. Project code: 201502A278.

References

He B., Fang W., Luo Q., 2008, Mars: a MapReduce framework on graphics processors, Proceedings of the 17th international conference on Parallel architectures and compilation techniques,ACM, 10, 260-269.

Hu R., Dou W., Liu J., 2014, ClubCF: A clustering-based collaborative filtering approach for big data application, IEEE Transactions on Emerging Topics in Computing, 2(3), 302-313.

Kuang G.F., 2012, The development of e-commerce recommendation system based on collaborative filtering, advanced engineering forum, 6(7),636-640.

Lee Y.H., Hu P.J.H., Cheng T.H., Hsieh Y.F., 2012, A cost-sensitive technique for positive-example learning supporting content-based product recommendations in B-to-C e-commerce, D ecision Support Systems, 53(1), 245-256. doi:10.1016/j.dss.2012.01.018.

Long S., Zhu W.H., 2012, Mining evolving association rules for e-business recommendation, Journal of Shanghai Jiaotong University (Science), 17(2),161-165.

Moon J., Chadee D., Tikoo S., 2008, Culture, product type, and price influences on consumer purchase intention to buy personalized products online, Journal of Business Research, 61(1), 31-39. doi:10.1016/j.jbusres.2006.05.012.

Papadimitriou S., Sun J., Disco, 2008, Distributed co-clustering with map-reduce: A case study towards petabyte-scale end-to-end mining, Data Mining, Eighth IEEE International Conference on. IEEE, 11,512-521.

Shekhar S., Gunturi V., Michael R., Evans, Yang K.S., 2012, Spatial big-data challenges intersecting mobility and cloud computing, MobiDE'12:Proceedings of the Eleventh ACM International Workshop on Data Engineering for Wireless and Mobile Access, 2, 1-6.

Xia X.W., Wang W.P., 2011, Collaborative Filtering recommendation algorithm based on trust model, Computer Engineering, 37, 26-28.

Yang S., Xue W., Xie Y.H., Wang X.Y., Zhu X.J., 2011, Collaborative filtering recommendation algorithm based on single-class classification, Computer Engineering, 37, 59-61.

Zhu Y., Su H.Y., Wang C.Q., 2014, Distributed collaborative filtering recommendation model based on expand-vector, Advanced Materials Research, 989, 2188-2191.

Zhu Y., Su H.Y., Wang C.Q, Yan B., Zheng H., 2014, Distributed collaborative filtering recommendation model based on expand-vector, Advanced Materials Research, 989, 2188-2191.