Advances in Social Science, Education and Humanities Research (ASSEHR), volume 68
2017 International Conference on Advanced Technologies Enhancing Education (ICAT2E 2017)

MCRS: A Course Recommendation System for MOOCs

Tao Huang[1], Gaoqiang Zhan[1,a] ,Hao Zhang[1,a*],and Heng Yang[1,a]
[1] National Engineering Research Center for E-Learning, Central China Normal University
[a]zhanghao@mail.ccnu.edu.cn [b]tmht@mail.ccnu.edu.cn

Keywords: MOOC, online course, big data, course recommendation, Apriori, Hadoop, Spark.

Abstract. With the popularization of MOOC platform, there is a tendency of big data in the number of online courses. Efficient and appropriate course recommendation can improve learning efficiency. According to the characteristics of MOOC platform, MCRS has made great improvement in course recommendation model and algorithm in this paper. The experimental results proves that MCRS's recommendation algorithm is more efficient than Hadoop Apriori algorithm, and recommend appropriate course to user. It turns out that MCRS is more suitable for MOOC platform.

1. Introduction

In 2012, the MOOC warmed up rapidly around the world. In complicated information space, user got lost due to information overload [1]. Most MOOC platform offer courses retrieving based on text and knowledge points, but these functions are not helpful in recommending courses according to user's own knowledge ability, learning methods and other dimensions [2]. Course recommendation system can analyze user-related data, and recommend appropriate courses to user, for example, using K-Means and Apriori algorithm to analyze learning situation of users, and recommend relevant learning materials [3].

Traditional course recommendation system should combine with big data framework, such as Hadoop, Spark. Wassan [4] proposed a data-mining model of association rules based on MapReduce, but it lacks further research, and pointed out that in order to improve the efficiency of educational data analysis, the construction of education of big data platform needs to rely on Hadoop, MongoDB, Cassandra and other tools in big data environment. West [5] mentioned that the education platform should have the ability of real-time analysis and feedback for student timely. It is significant to have a research of the course recommendation algorithm based on distributed computing framework.

Mentioned in paper [3], the Apriori algorithm is based on a closed educational environment. In paper [6-8], the related recommendation algorithms are based on distributed computing environment in MOOC environment. The course recommendation algorithm is based on Apriori algorithm and Spark in this paper. The analysis process mainly includes two parts: mining 1-frequent item set as the initial candidate data set, and then mining k-frequent item set according to the support until not satisfied, the results of analysis are stored in database as course recommendation rules.

2. MOOC oriented course recommendation model

For information consumer and producers, it is difficult to locate the satisfied information accurately and push their information to the user who need it. Recommendation system is helpful to solve the problem, the evaluation dimension of recommendation system includes three aspects: user dimension, item dimension, user preference dimension. Recommendation model deals with the related data, and recommends appropriate items to the appropriate user.

MOOC oriented course recommendation system, the number of courses presented as PB-scale. In the process of transplanting traditional course recommendation model into MOOC, big data framework should be included. We propose the MCRS based on above analysis, which is a MOOC oriented course recommendation system. It consists of three parts:

1) Data collection and preprocessing, importing students' course enrollment information into HDFS through Sqoop, and preprocessing the data on Hadoop, storing the simplified data;

2) Association rules mining, Spark can directly read HDFS data into memory for calculation, combined with the advantages of both, it is easy to analyze the association rules of formatted data.

3) Recommending course according to the parameters, filtering the results according to confidence, and exporting the results to relational database, if there are students in accordance with the rules, and the course would be recommended.

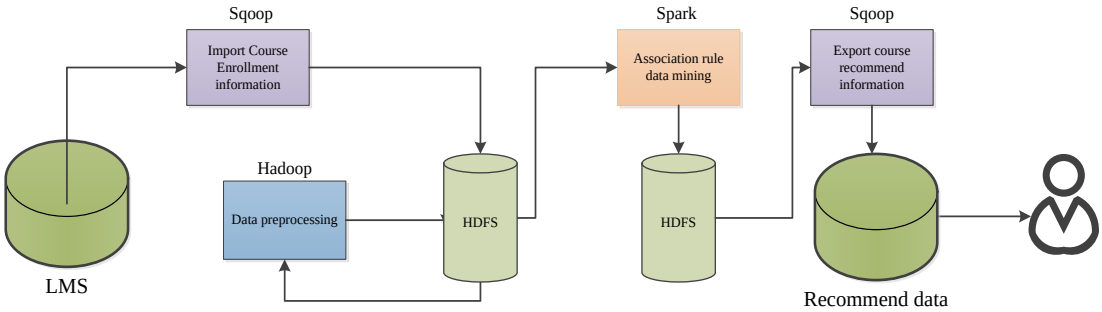

Fig. 1 Course recommendation model

3. Algorithm Analysis

Association rules mining is a theory of collaborative filtering algorithm, such as Apriori. The basic idea is a recursive algorithm based on the idea of two-phase frequent item sets. The purpose of association rules is to find the relationship among items in a data set, also known as Market Basket analysis. Firstly, scanning the data set and storing as the original data set, and then calculating the number of same item. Obtaining 1-Frequent item set by pruning according to support, Obtaining 2-candidate item set from the permutation and combination of 1-frequent item set's keys, Obtaining 2-frequent item set by pruning 2-candidate item set according to the support, based on the recursive analogy, calculating the K-frequent item sets.

The first phase is a simple word count process that calculates 1-frequent item set according to the support, calculating frequent item sets as original data for calculation, as shown in Fig. 2.

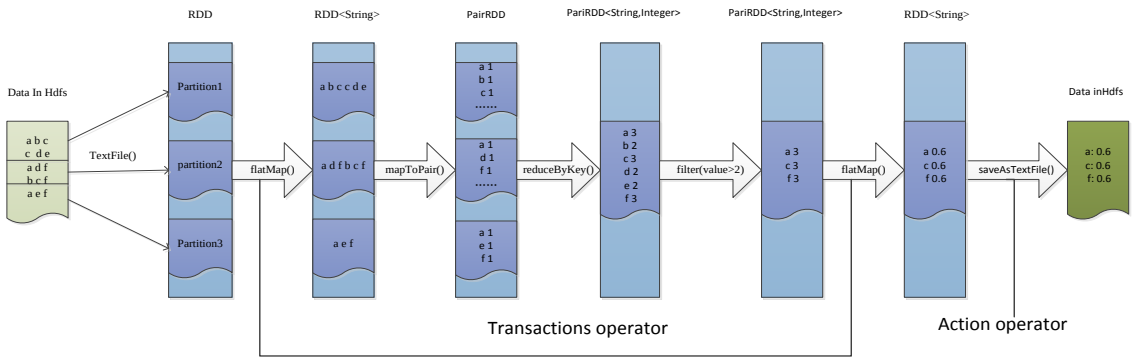

Fig. 2 Mining 1-frequent item sets

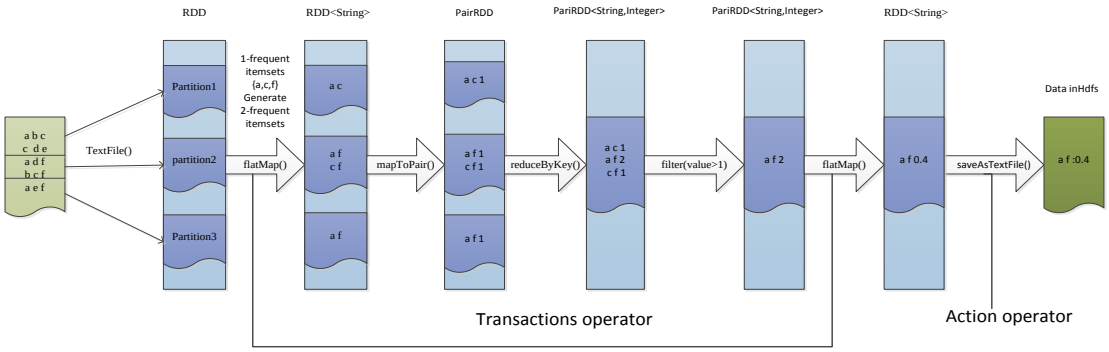

Fig. 3 Mining 2-frequent item sets

The second phase is to iterate the original data, and output all K -frequent item set that satisfied the support. k-frequent item set add the process of mining multi-item set in map process, each map process in accordance with the K item set , as shown in Fig. 3.

4. Experimental Analysis

4.1 Data Set. Three benchmark datasets are used during the experiment, IBM provided the dataset for Apriori algorithm, T10I4D100K, T25I10D10K (generated by a random iteration of the database), Harvard and MIT published the edX learning data in 2012-2013, HMXPC13.csv.

4.2 Algorithm efficiency. The minimum support is 100/transaction, the maximum frequent item set is 8, each experiment was carried out three times to reduce the deviation of the experiment. Collecting the results of the experiment, recording the time cost for each experiment and mining the k-item set. The left side of Fig. 4 shows the experimental results of T25I10D10K data set. The difference in the small number of data is not obvious. The right side shows the experimental results of T10I4D100K, the efficiency increases from two to four times. The advantages of Spark will become more obvious in large clusters, the experiment results show that the efficiency of association rules mining algorithm proposed in this paper is higher than Hadoop based association rule mining algorithm.

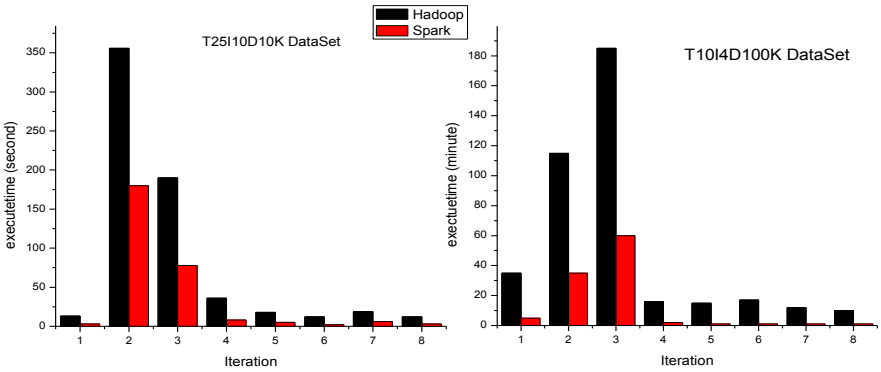

Fig. 4 Execute time on Spark and Hadoop

4.3 Course Recommendation Experiment. Table 1 is the experimental results of HMXPC1. The experiment takes 100,000 data as the training data to analyze the rules between courses, the remaining 19024 used for testing. According to the support of the experimental results, and calculating the confidence between the courses, rule A-> C: Confidence = Support (A, C}) / Support ({A}). Calculating the rate of course enrollment for the recommended course in the predicted course enrollment data, according to the highest confidence of the course rules. Recommending course 2 to those who enrolls the course 1 and 3, recommending course 3 to those who enrolls the course 1 and 2, and then verify the efficiency of the rules with 19024 historical data. The confidence of course 2 after courses 1 and 3 is 0.4928, there are 284 users enroll courses 1 and 3 in the historical dataset, 60 users among them enroll course 2 in subsequent courses, the efficiency of course recommendation is 26.6%. Thus, the MCRS can recommend the appropriate course for some users.

Table 1 Confidence and Recommend Rate

Course Rule	Confidence	Recommend Rate
1,3->2	0.4928	26.6%
1,2->3	0.4632	25.2%
3->1,2	0.1257	Not recommend
3->2	0.4036	24.4%
1->3	0.3983	25.3%
2->5	0.0213	Not recommend

Summary

MCRS is a course recommendation system that satisfies the demands of today's big data environment, and the basic algorithm is based on Apriori algorithm and Spark mainly.

Based on the MCRS, we can implement different algorithms on the basic data mining algorithms, and use Spark to implement the algorithms such as genetic algorithm, collaborative filtering, decision tree. As proposed in introduction, it improves the efficiency of course recommendation. To promote personalized learning, carrying a lot of education data mining work on MOOC platform. In addition, it needs to explore on how to use Deep Learning to EDM to design better and smarter education system, and how to make recommendation accurately and efficiently.

Acknowledgements

This work was supported by the National Office for Education Sciences Planning Project No. CCA140152.

References

[1] Rendle, Steffen. "Factorization machines with libfm." ACM Transactions on Intelligent Systems and Technology (TIST) 3.3 (2012), 57.
[2] Gaikwad,T. and Potey. Personalized Course Retrieval Using Literature Based Method in e-Learning System., IEEE Fifth International Conference on Technology for Education (2013).
[3] Aher, S.B. and L.M. Lobo, Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data. Knowledge-Based Systems, 51(2013), 1-14.
[4] Wassan, J.T., Discovering Big Data Modelling for Educational World. Procedia - Social and Behavioral Sciences, 176(2015), 642-649.
[5] West, D.M., Big Data for Education: Data Mining, Data Analytics, and Web Dashboards. Hrvatska znanstvena bibliografija i MZOS-Svibor. (2012)11.
[6] Zhou Q, M.C.Y.D., Research progress on educational data mining: A survey. Ruan Jian Xue Bao/Journal of Software, (2015)26, 11, 3026-3042. http://www.jos.org.cn/1000-9825/4887.htm.
[7] Wang, Y., M.H. Tseng and H.C. Liao, Data mining for adaptive learning sequence in English language instruction. Expert Systems with Applications, ,36(4) (2009) 7681-7686.
[8] Salehi, M., I.N. Kamalabadi and M.B.G. Ghoushchi, An effective recommendation framework for personal learning environments using a learner preference tree and a GA. IEEE Transactions on Learning Technologies, 6(4) (2013)350-363.