

# Research on e-commerce intelligent service based on Data Mining

Jianxi Zhang <sup>1,2</sup>, Changfeng Zhang <sup>2</sup>, and Huaiizhi Yu <sup>1</sup>

<sup>1</sup> Computational Sciences, Shandong Province Key Laboratory of Storage and Transportation Technology of Agricultural Products, Ji'nan, Shandong, China

<sup>2</sup> Food Sciences, Shandong Institute of Commerce and Technology, Ji'nan, Shandong, China

**Abstract.** With the rapid development of electronic commerce in China, a large amount of information data will be generated at every moment. How to excavate useful information is becoming an important problem in the big data age. Firstly, the smart service model of E-Commerce based on data mining was proposed, and user group mining, user interest mining, industry and domain knowledge mining and business association mining were used to bridge the gap between the big data application and requirements of smart service. Then the technical support system of E-Commerce data mining based on Hadoop platform was suggested to provide technical solution for implementation of smart service applications. And finally, the scenario knowledge recommendation service with the support of big data mining were discussed.

## 1 Introduction

In recent years, with the development of Internet, there is an e-commerce fever all over the world. At all times, the e-commerce system produces a large amount of information data which has long been characterized by large data. With the advent of the big data age, seeking effective processing technology and methods to mine useful information has become an urgent demand in the real world. On the one hand, the electricity supplier website can retain users effectively and improve customer loyalty as users get rid of the plaguing of large amounts of goods and enjoy a better shopping experience. On the other hand, the information recommendation system can recommend to users the goods they are not familiar with but are very fond of, prompting users to buy new goods, so as to improve the overall sales of e-commerce sites.

The research of e-commerce recommendation system started earlier in foreign countries. IBM's A. Ballman et al. developed the SpeedTracer system for data mining and analysis based on Web log [1]. Paolo Buon et al. proposed user information (called "explicit information") and user behavior (called "implicit information") as input to the recommendation system [2]. Mehmet H. Goker of Stanford university and others have come up with a conversational recommendation system to help users filter information [3]. Kwong Hiu Yun and others have used

a variety of data mining techniques to establish an online book recommendation system [4]. J. ben Schafer, at the University of Minnesota, and others have proposed the use of collaborative filtering to produce recommended related technologies [5].

The development of the e-commerce in our country started relatively late. The research results are relatively backward compared with the foreign countries. Deng Ailin et al. proposed a collaborative filtering Item-Based algorithm [6] [7]. Xiong Xin has improved the collaborative filtering algorithm in the personalized recommendation system and introduced the concept of stratification [8]. Wu Xizhi and others put forward the association rules based on the knowledge base and the profit of the goods [9]. In the study of personalized recommendation system, Li Feng and others proposed a new personalized recommendation algorithm based on the characteristics of the goods [10].

## 2 E-commerce intelligent service model based on Data Mining

Considering the current development of e-commerce, the core business mode of e-commerce intelligence service basically includes four aspects. (1) Intelligent group knowledge sharing that impel knowledge to transform, transfer and broad as to make knowledge easy to learn. (2) Intelligent knowledge recommendation that is according to users' interest

\* Corresponding author: [author@e-mail.org](mailto:author@e-mail.org)

and requirement, as to provide ubiquitous, individualized and scenarized knowledge service, to make knowledge easy to utilize.(3)Intelligent knowledge navigation that provide knowledge topic, knowledge hot point, knowledge development trend, knowledge relation and knowledge extensibility for users to make knowledge easy to understand.(4)Intelligent business optimization that take the user's needs as the guidance, carry out the activities of reading popularization and lecture.

Internet technology makes the interconnection and collaboration between the resources of the e-commerce platform, the users, the users and the resources to reach the unprecedented breadth and depth. In particular, with the diversification of the use of terminals and the wide use of social tools, a full and stereoscopic large electronic business data has been formed. Important large data resources, which can be used to meet the core needs of intelligent services, include three major categories.(1)User data: User behavior data(Including explicit data and implicit data),Terminal aware data and Social data, etc.(2) Knowledge resource data.(3) Business process data. Its composition is shown in Table 1.

**Table 1.** Major data resources in E-commerce

Data type		Data composition
User data	Explicit data	Purchase, Evaluate, Score etc.
	Implicit data	Browse, Collection, Click stream etc.
	Personal information	Gender, Age, Occupation etc.
	Social data	WeChat, Mailbox, Cell-phone number etc.
	Terminal aware data	Location, Time, Scene, Device parameters etc.
Knowledge resource data		Commodity classification, business shop etc.
Business process data		Advertisement, Search, Commodity purchase, Feedback data, etc

Data mining is one of the key technologies to make large electronic business data play a role. This article presents an electronic business intelligence service model based on large data mining. It uses large data mining technology as a bridge to communicate the large data application of e-commerce and the demand for intelligent services.

(1) User group mining

Mining user group to realize group knowledge sharing. Build large-scale social networks based on basic data such as personal work experience, work field, and social data such as WeChat, Cell-phone number and mailbox. It can apply classification, clustering, frequent patterns mining and other mining

methods to mining user communities or key persons, and study the transformation of implicit data into explicit data, transfer of knowledge and dissemination, so as to realize user group knowledge sharing.

(2) User interest mining

Mining user interest to realize personalized, scene,and ubiquitous knowledge recommendation and push. Analyze user explicit and implicit data, as well as mobile phone, tablet and other use terminal perception data, mining user deep demand. According to the special scene, it can recommend various resources to users in different fields, layers and stages, and realize intelligent recommendation and push.

(3) Industry and domain knowledge mining

Mining industry and domain knowledge to realize automatic business recommendation. Mining subject and topic association based on keywords, establish knowledge semantic network in industry and domain to obtain the industry knowledge hot spot, combining common word analysis and clustering analysis method. At the same time, adding time parameter can show the dynamic change and development direction of the industry development.

(4) Business association mining

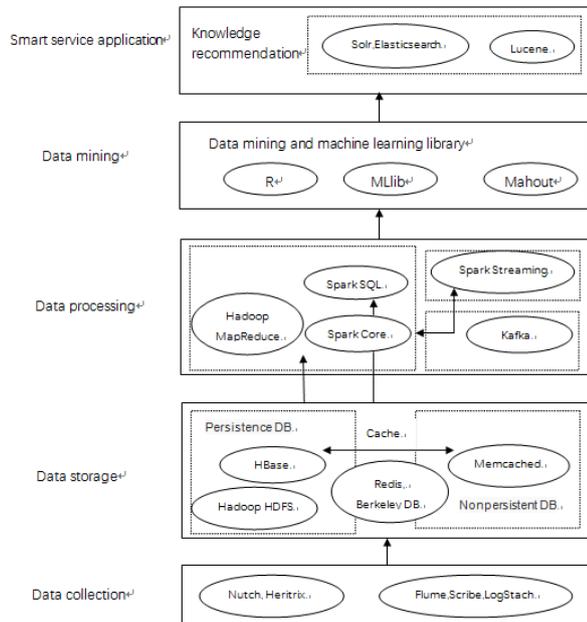
Mining business association to achieve business optimization. Analysis of association rules for management data such as advertising data, commodity data, retrieval data etc. to find various associations related to user requirements such as a period of time, the association of a certain type of user with a certain business requirement, etc. Optimizing the business process of e-commerce and analysis of the association rules between the circulation data and the external data. Discover the association between the number of visitors and the location, a time node, or the association of an event with the number of visitors, which provides support for the electronic platform to carry out advertising promotion, commodity recommendation and other service activities.

### 3 Large data mining technology system for intelligent service

#### 3.1 The technical support system based on Hadoop

The Hadoop platform is an open source platform produced after Google's MapReduce model [11],It has the characteristics of scalable, high efficiency, low cost and so on. Smart service requires high real-time. In addition to large static data, large data mining of electronic commerce for intelligent service needs to

emphasize the dynamic data in order to carry out real-time data analysis and dynamic integration to find valuable knowledge. In this paper, a large data mining technology support system based on Hadoop is constructed to support the real-time calculation and mining of large data of e-commerce. as shown in Fig. 1.



**Fig. 1.** Data mining technology system for intelligent service.

### 3.2 Data collection, storage and processing

The bottom layer is e-commerce data collection in Fig.1. The information of external industry, industry news and other information can be collected from the Internet by the open source web crawler system such as Nutch and Heritrix [12]. The internal data of e-commerce can be collected through the Flume system provided by Cloudera. Flume is an open source distributed mass log collection system. It is safe and reliable. The user's access log can be periodically transmitted and saved to a distributed store for subsequent tracing and analysis.

In the data storage layer in figure 1, Hadoop's HDFS provides the most basic persistent distributed file system. HBase and Mongo DB provide functions like relational databases for advanced application development. The column storage of HBase is convenient for data definition to be changed at any time, and is suitable for storage and query of large-scale data. Mongo DB supports complex hierarchical structures that provide greater flexibility for storing large data with 1 non-standard social text. Redis, Berkeley DB and Memcached provide a caching mechanism for HBase and Mongo DB databases, which greatly improves the system response speed and reduces the pressure of persistent storage.

In figure 1's e-commerce data processing layer, Hadoop's MapReduce and Spark Core Core components are designed for batch processing which can be used to analyze and operate mass data with the idea of mapping and specification. For example, they can count the hot key words generated by recent users' retrieval of goods. Spark SQL combines different formats and structured data from multiple data sources, providing a shortcut for users who are familiar with relational SQL languages. Using the Kafka message mechanism, the change of data can be pushed to each data processing system for incremental updating in time. Spark Streaming provides a flow computing framework based on the idea of mapping and specification to further improve the real-time performance of the processing.

### 3.3 Data mining algorithm and application

In the era of big data, data mining faces new challenges. Such as, when a traditional theoretical model meets massive data, a single machine can't cope with it. The Hadoop based data computing framework provides a solution for its distribution implementation.

**Table 2.** Data mining algorithm and application in MLib Library

	Data mining algorithm	Application
Classification and regression	Linear regression, logistic regression, Isotonic regression, Logistic regression, Bias-classification [13-15], Decision tree classification, BP, RBF neural-network, SVM [16]	User classification; Commodity resources classification; Time, location and other environmental situation interest prediction; Age, sex, occupation and other user contextual interest prediction
Clustering	K-Means Clustering, LDA [17]	Keywords cluster analysis; Hot topic mining; User community and key people mining
Recommend	Collaborative filtering recommendation, ALS Recommendation [18]	Knowledge recommendation, including user based recommendation and resource based recommendation
Association rules	FPGrowth [19]	Business data association analysis; Knowledge recommendation

Deep learning	DBN CNN	Knowledge map deep learning; Content feature mining of users and resources
---------------	------------	---

In the data mining layer of Figure 1, MLlib, Mahout, and R are all data statistics, mining and analysis software that can run on the Hadoop platform. MLlib is an extensible data mining and machine learning library in Spark. It not only includes classification, regression, clustering, collaborative filtering and other traditional algorithms, but also incorporates a new advanced learning algorithm. Table 2 lists the main data mining algorithms in the MLlib library, as well as their applications in large data analysis of e-commerce intelligence services.

In the smart service application layer shown in figure 1, Lucene is an open source full-text search engine toolkit proposed by Apache. Solr and Elasticsearch are two search servers based on Lucene, which can provide the basis for the application of retrieval, recommendation and knowledge q&a. The data mining acquired data wisdom into various service applications can provide high quality and intelligent service for users.

#### 4 Discussion on smart services supported by large data mining

The application of large data mining technology to provide users with intelligent service is the development trend of e-commerce services. This paper makes a brief discussion on the scene based knowledge recommendation intelligent service under the support of data mining.

The scene-based information recommendation service recommends the commodity resources to the user based on the special scene that the user is currently in. However, it is necessary to use all contextual information related to human-computer interaction to demarcate the current special scene. The application of smart mobile terminal, such as smart phone and tablet, not only provides convenience for users, but also provides rich situational information for scene based knowledge recommendation. Real-time perception and mining of mobile situational data can provide real-time dynamic personalized recommendation for users, make knowledge resource recommendation fit with users' scenes, better meet users' needs, and make knowledge easy to use. Time and location are two important mobile situational information, which can be collected by a variety of sensors, such as GPS, WiFi, Bluetooth and so on. The key of personalized recommendation in mobile

context is user behavior pattern mining. Through classification and regression algorithm, it reveals user preferences and rules of life, and improves recommendation efficiency.

#### 5 Summary

Application of big data mining technology to implement intelligent, personalized and initiative smart service and further promote business mode innovation is the inevitable trend of e-commerce service development. The electronic commerce data mining technology support system based on Hadoop platform can complete data collection, storage and processing, and realize the real-time mining of large data. With the support of large data mining, the intelligent service, such as the scene based information recommendation service, has become a new mode of e-business service. The study of this paper has a certain reference to the application of large data mining methods and technology to realize intelligent knowledge service in e-commerce.

#### Acknowledgment

This study was supported by The National Key Research and Development Program of China (2016YFD0400900).

#### References

1. A.Ballman, Wu .K.L, Yu.P.S. A Web usage mining and analysis tool. IBM Systems Journal. 1998,89-105.
2. P.buono. Integrating User Data and Collaborative Filter in a Web Recommendation System [J]. Third Workshop on Adaptive Hypertext and Hypermedia,2001,7(3):129-140.
3. Mehmet H.Goker, Cynthia A.Thompson. A Conversational Recommendation System, DOI= Http://www.agr.informatik.unikl.de/gwcb2k/goker\_thom pson.pdf.
4. Kwong H Y. Recommendation System for Large Online Book seller. DOI=Http://www.Stanford.edu/khyung/recommender\_sy stem.pdf.
5. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. Collaborative Filtering Recommender System.
6. Deng A L, Zhu Y Y, Shi B L. Collaborative filtering recommendation algorithm based on project score prediction [J]. Journal of software, 2003 (09):64-66.
7. Deng A L, Zuo Z Z, Zhu Y Y. Collaborative filtering recommendation algorithm based on project clustering. Minicomputer system,2004,09(21).
8. Xiong X, Wang W P, Ye Y X. Personalized recommendation algorithm based on concept stratification [J]. Computer application, 2005(05):89-90.
9. Wu X Z, Run J, Su L M. Application of Individualized Recommendation Based on association rules in

- traditional business [J]. Lanzhou Academic Journal,2007(02):55-56.
10. Li F, Li J H, Wang R L. Personalized recommendation algorithm based on commodity characteristics [J]. Computer engineering and Application,2007(17):89-91.
  11. Lin,Zhang L, He Q, et al. Paralled Implementation of Apriori Algorithm Based on MapReduce [C]//The International Conference on Business Computing and Global Informatization. Washington,DC,USA,2011;475-478.
  12. Chen X G. Research and Realization of E-commerce Monitor System Based on Focused Web Crawler[J]. Information Technology Journal, 2013, 12(18).
  13. HARRY Z, SHENG S L. Learning weighted naive Bayesian with accurate ranking[A].Washington,DC:IEEE,2004.567-570.
  14. Friedlman N, Geige D, GoldszmidtM. Bayesian Network Classifiers. Machine Learning,vol.29,pp131-163,1997.
  15. Goldenberg A, Moore A. Tractable learning of large Bayes net structures from sparse data[C]//Proceedings of the twenty-first international conference on Machine learning. ACM, 2004: 44.
  16. Fan J, Zhang X K, Tian X, et al. An Adaptive Topic Crawler for Electronic Public Opinion[J]. Advanced Materials Research, 2013, 765: 1451-1455.
  17. Salton G, Wong A, Yang C S. A vector space model for automatic indexing[J]. Communications of the ACM, 1975, 18(11): 613-620.
  18. Mercer R L. Class-Based n-gram Models of Natural Language[J]. Computational Linguistics, 1992, 18(4):18-4.
  19. Guyon I, Elisseeff A. An Introduction of Variable and Feature Selection[J]. Journal of Machine Learning Research, 2003, 3:1157-1182.