Advanced Science and Technology Letters
Vol.134 (CIA 2016), pp.75-81
http://dx.doi.org/10.14257/astl.2016.134.13

Research on Personal Business Information Service based on Big Data

[1]Liu Hongyan, [2]Liu Zhenyu

[1]Beijing Information Technology College
1241420132@qq.com
[2]National Center For the Perform Arts
112858658@qq.com

Abstract. With the development of the times, e-commerce has affected people's lives in all aspects. The user's demand for e-commerce is more targeted, which makes the personal information service system based on the big data become a topic of increasing concern. Through the establishment of the personal information recommendation service pattern based on big data and analysis the factors affecting the user's decision, we can meet the needs of users, make our groups stable, improve service quality, and illustrate the feasibility and effectiveness of the model.

Keywords: big data; cloud computing; cloud platform; Resource sharing.

1 Introduction

Along with the era of big data is the quality of data utilization, the test of controlling the data behavior, and requires people to have the ability to have a corresponding improvement. For enterprises, the big data can do the data support for recommendation service. The concept of electronic commerce in the era of big data produces fundamental change, also produces a new mode of information management, e-commerce industry service innovation can take advantage of this change, better use of data information.

How to provide customers with the required information and improve the accuracy of the information to get the customer satisfaction is a urgent problem needs to be solved. Based on the analysis of customers, this article puts forward the analysis of the relevant customer data to identify the customer characteristics, so as to provide a solution for the different customers with personalized information services [1-3].

ISSN: 2287-1233 ASTL

2 Related research

2.1 Research on business behavior theory

Consumer behavior research began in the end of nineteenth Century, and has been rapidly developed since 1960s.The main research framework of consumer behavior can be simply summarized as consumer purchase decision-making process and the influencing factors. In the study of consumer behavior, social psychology professor Ajzen and Fishbein's theory of rational behavior is one of the most extensive range of theory. Research shows that TRA has successfully predicted the behavior of consumers in many areas. According to the TRA theory, consumers can make rational according to the results of their own interests, but also to meet the needs of other people's behavior way to implement, and pointed out that the behavior motivation and access to information on consumer behavior. Beckett et al. Point out how consumers make choices to a large extent can be controlled. When consumers are in a higher degree of uncertainty and the level of the required input is also relatively high, they are more difficult to be completely rational and positive decisions, and more to reduce the uncertainty. "Postmodern" producers are no longer trying to dominate and control the consumer, but to change the concept of the framework, to establish the relationship between producers and consumers, and to control and influence of consumer's decision-making behavior through the "connection" management technology.

So it has been difficult to systematically manage all the impact mechanisms. Along with the continuous development of information technology and the further expansion of the market scale, business model management and the operation of social systems are more and more integrated with information technology, online consumer decision-making behavior has also been increasing and continued to deepen.

2.2 Personal business recommendation system

With the rapid development of network information technology, and personalized recommendation for mass products and services has led to the transformation of marketing and business model, more and more websites use recommendation system to provide online consumer products recommended service. The rapid development of the Internet makes consumers have the possibility of unlimited choices. We have left the information age and entered the era of recommendation. The collection of information is no longer a problem, but the key is how to make the right decisions based on information. The value of measuring and analyzing the behavior of the public is realized by more and more companies. Through information technology and software, we can observe the behavior of consumers and give guidance. Recommended services can not only help customers find the goods you want, but also to some extent, affect the customer's choice. The definition of the proposed system for Resnick and Varian is currently the most cited: "Recommendation system is the use of the website system to provide consumers with information and related suggestions,

simulation of the traditional way of sales staff to help consumers complete the purchase process, to help consumers make purchase decisions [4-6].

Since the seventy's of last century, the research of personalized recommendation service has begun to start and collaborative filtering algorithm is also proposed. During the twenty years of research, the development of personalized recommendation system has been basically formed in theory, but there is still a relatively slow development in the distance. The research results of Rose and Lees show that the time of searching information on the Internet is a key factor to affect the sales of goods. The time cost is related to site navigation, product search function and the whole website design. By studying the user's reaction to the shopping site recommended to their own unfamiliar products, Cooke et al. believe that acceptance and acceptance are derived from the results of personalized recommendation. In 2011, Junyean proposed the hypothesis and the model of personalized recommendation service from the cultural dimension and carried on empirical researches.

3 The proposed scheme

Individual factors that affect the user's purchase of personalized products are mainly user's income level and online shopping experience. According to Gillet's research, the people whose average income level is higher than the average level of the community are more likely to buy online. Generally, the high income group for the risk of tolerance is also higher than ordinary people. At the same time, the higher the income level of the crowd in the online shopping overhead is higher than others, as well as the number of shopping. Personalized products tend to recommend the user's preferences of goods. Sometimes user generated buying behavior is simply because of the love and takes the net purchase as a way of pleasure. The difference in the level of user income will affect the attitude of shopping for personalized recommendation. The proposed scheme is based on the following assumptions: 1. The higher income level of the users who have online shopping behavior is, the more likely for users to buy the goods of personalized recommendation. For the online shopping experience, this scheme has the following hypothesis, experience is the accumulation of past knowledge and the objective results generated in the participation behavior. Users in the process of online shopping information collection process and purchase decision-making stage will be influenced by the factors of their past experience. The goods personalized recommendation is the goods that sellers recommended initiatively to the users, which requires a higher intention of users. 2. The richer the online shopping experience is, the stronger the intention to purchase personalized recommendation of goods.

Compared with shopping offline, online shopping has many advantages. But the net purchase of the most serious flaw is that the user cannot really feel the good or bad things. Empirical type of goods refers to the goods that users have purchased. Search products are the goods that users select after searching and comparing. Personalized recommendation is based on the web site of the user to browse the web and the shopping process to leave the data for the user recommended products. So the relevant degree of the recommended product and the user's ideal commodity will

directly lead that the user whether the purchase intention of this commodity. Therefore, it comes 5th assumption: The degree of correlation between the recommended items and the ideal product will affect the user's shopping intention [7-8].

Popularity is mainly refers to the user's understanding of a shopping site, which measures the size of its reputation in the objective scale. Users trust the site higher, the intention of its online shopping behavior is more obvious. For the user, the higher the site's visibility is, the higher the degree of personalized recommendation of goods and ideals are, as well as reliable. Therefore, the 6th assumption is that the visibility of the site will affect the users to buy personalized recommended product behavior.

3.1 Research on Business Consumption Behavior

Deep belief network is a generative model proposed by Hinton Geoffrey of University of Toronto in 2006. By training the weights between the neurons in the model, the entire depth neural network generates training data in accordance with the maximum probability. Deep neural network algorithm has high efficiency, fast convergence rate and the accuracy of analysis. However, the construction of deep neural network is complex, and it is difficult to judge whether it is optimal or not, so it is suitable for large data. So it is proposed that the predicting Personal business behavior of deep neural networks based on Cloud Computing method suitable for large data. Deep learning technology's advantages are mainly in its ability to abstract generalization of data, it can express the learning objects as a high order, so that feature matching, pattern recognition, text processing and other aspects of the problem is effectively solved. These problems are involved in the process of providing knowledge service for scientific instrument development. In the process of analysis, processing and integration of all kinds of data resources, mainly contains the following technology: Chinese word segmentation, image the identification, classification, form, language model, text semantic analysis classification, etc. Because of the huge amount of data and the complex of knowledge structure, the method of manual intervention is time-consuming, so it is necessary to deal with the information resource. Faced with such complex data characteristics, the traditional machine learning lacks the ability to express data features. Once the data complexity is improved, the dimension of disaster is likely to occur. Traditional machine learning methods also lack the ability to performance of multidimensional nonlinear problems. The network framework of deep learning is composed of multi-layer nonlinear computing model, which has the ability to describe the complex object, and it can be used in the distributed mode, which can improve the learning efficiency of the model. Through the experimental verification of the literature, it shows the learning performance of the greedy layer by layer unsupervised deep learning method obviously. Since deep learning has these advantages, it has been widely used in the various problems mentioned above. In the process of providing knowledge services, it is particularly focused on the depth of information resources and the related knowledge. And using deep learning technology in the process of knowledge and information processing, can significantly improve the accuracy and efficiency of data processing, and also enhance the support effect of knowledge service to the development, production and use of information.

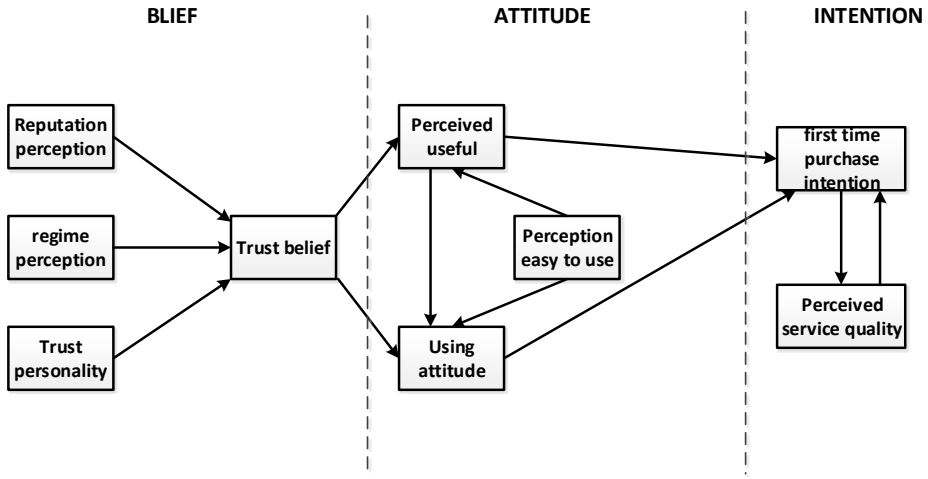

Fig. 1. forecasting model

According to this model, we can predict the personal business information intention, but for this hypothesis testing, BP algorithm for MapReduce, the BP algorithm split into two parts：In the first Map process, the model of batch training is used to function on the network, and the network weights are performed after a certain number of iterations. The second part Reduce combines the output of the Map to obtain a new network weights, and determine whether the adjustment of the iteration according to the weight of the network. The following is a concrete implementation approach and pseudo code. In the Map function, the network read the network weights from the file system, initialization network, read the sample. After a certain number of training has been achieved in the node (The condition may be a certain number of iterations, the output error may be up to a certain level, the weight change amount reaches a certain level. In this study, we use the output error reaches a value (0.01)) the partial sample can be considered as the convergence state of the node. Map of the output is a team of value pairs. After Reduce obtains the value pairs, process it. Define an implementation Writable interface class WeightWritable as a value type of the Map. WeightWritable implements the Writable Hadoop interface, which conforms to the standard of Hadoop serialization, and records the weights of the network after training. Each turn of the learning result is sent to a Reduce to be integrated, so the output of the type LongWritable key is set to a uniform value (0).

The process of the Map function is as follows:

1. Read the HDFS's weights records on the preservation of the network.
2. According to the network weights read, instantiate a neural network.
3. Process Text and extract the input of the training sample and the target output. Then, Normalized.
4. Use a sample to train, until the end of a training condition is met.
5. Get the weights in the network and instantiate WeightWriteble.
6. Output WeightWritable

Function Reduce receive the output of Mapcalled <1|01^Writable, WeightWritab10 as its own input and the output is <LongWritable, IntWritab10. Where value is the type of int and means whether the network needs to be carried out in the next iteration. The value of 1 is required to reiteration and the end of training for the 0. Reduce counts the total weight of each map and calculated the average value as the weight of the whole network. After calculating the weight of the arithmetic average (The arithmetic mean value is the addition of each weight matrix, and then the division operation. WeightWritable realizes the accumulation function and division function.), function reduce can read the network weights in file system saved before training and compare two weight records. When the difference between them is less than a certain level, the output of the int sets 0, otherwise sets 1. Using the weights WeightWritable of the value after training update the weight file in file system and the initial weights are used as the initial weights of the next iteration.

4 Experimental results and analysis.

This experimental environment uses 4 ordinary personal computer and the configuration is as follows:

NameNode: one computer; CPU is 2.3 GHz; Memory is 2G; Hard disk is 300G;

DataNode: 3computers; CPU is 2.67 GHz; Memory is 1G; Hard disk is 80G;

Software environment–Ubuntu10.10; JDK 1.6; Hadoop 0.20.1; HBase 0.90.3.

There are three purposes in this experiment: 1.verify the feasibility of the neural network algorithm based on cloud platform; 2.Test the convergence rate of the algorithm;3.Test the training accuracy of the algorithm。

The data set used in this experiment is derived from the Data Set of UCI named Breast CancerWisconsin (Original).The data set is a collection of typical models for pattern classification, which can be used to test the classification results of classification algorithms. The data set contains 699 sample data. 16 of them are missing. The first 450 data are used as training samples. The remaining 233 are used as test data. Each record contains 11 domains, where the data field 1 is the 1strecord of the ID. The data field 2-10 is the data input mode and the data field 11 is the record of the classification. Each record is divided into two types, positive one and negative one. In the experiment, the training sample is divided into several parts to carry on the training to the multiple Mapper. The network classification accuracy test is carried out in the experimental drive function. After each experiment run for10 times, then calculate the average value.

Table 1. The time required to predict

File（M）	7.12	17.8	28.5
Time 1	2612	6645	10563
Time 2	2702	6650	10569
Time 3	2628	6570	10650
Time 4	2630	6518	10578

Supervised learning training number v.s. average error

The number of training	Average Error
8	125%
25	110%
520	52.6%
1200	5.6%

5 Conclusion

In this paper, according to the data characteristics of the personal business service information, a large data processing method is put forward. This method takes the characteristics of large data into account that large data is large and need to dispatch, and according to the characteristics of large data, we propose a large data analysis platform based on cloud computing platform. This method can effectively improve the prediction accuracy of the system, and with the increase of the number of training, its accuracy is significantly increased, at the same time, it has self - learning habits.

References

1. Ji, C., Li, Y., Qiu, W.: Big data processing in cloud computing environments[C]//Pervasive Systems, Algorithms and Networks (ISPAN), 2012 12th International Symposium on. IEEE, 2012: 17-23.
2. Fox, A., Griffith, R., Joseph, A., Above the clouds: A Berkeley view of cloud computing [J]. Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS, 2009, 28: 13.
3. Gantz, J., Reinsel, D.: The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east [J]. IDC iView: IDC Analyze the Future, 2012, 2007: 1-16.
4. Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-the-art and research challenges [J]. Journal of internet services and applications, 2010, 1(1): 7-18.
5. Ostermann, S., Iosup, A., Yigitbasi, N.: A performance analysis of EC2 cloud computing services for scientific computing [M]//Cloud computing. Springer Berlin Heidelberg, 2010: 115-131.
6. Marx, V.: Biology: The big challenges of big data [J]. Nature, 2013, 498(7453): 255-260.
7. Shekhar, S., Gunturi, V., Evans, MR.: Spatial big-data challenges intersecting mobility and cloud computing[C]//Proceedings of the Eleventh ACM International Workshop on Data Engineering for Wireless and Mobile Access. ACM, 2012: 1-6.
8. Chen, H., Chiang, RHL., Storey, VC.: Business Intelligence and Analytics: From Big Data to Big Impact[J]. MIS quarterly, 2012, 36(4): 1165-1188.