Transactions on Computational and Scientific Methods | Vol. 2, No. 1, 2022
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Analyzing the Effectiveness of Big Data-Driven Information Recommendation Strategies

Isabella Hernandez

Department of Computer Science, University of North Texas, Denton, USA

Isabella.h9032@unt.edu

Abstract: As time progresses, the recommendation of information based on big data has increasingly become a significant research task. This paper categorizes users into ordinary and advanced users. It then proposes a big data-based information recommendation strategy. The experimental findings indicate that the proposed strategy is effective.

Keywords: The information recommendation, Big data.

1. Introduction

With the rapid development of today's information, the information recommendation is facing new challenges. Chinese Academy of Engineering, Chinese Academy of Sciences Institute of Computing Technology, chief scientist GuoJie Li pointed out that the big data will become the new focus of information technology, and form to a new interdisciplinary: scientific network data [1]. Meanwhile, due to the current organization of information technology in the construction of knowledge service, knowledge creation, knowledge organization and knowledge storage and so has a wide range of applications, making social knowledge of the service requirements more demanding, the potential of knowledge demand for value-added services mining, knowledge assessment, data analysis, have already begun to appear or has resorted to the expression in everyday behavior among readers. In this paper, the use based on big data technologies, in the understanding of an overview and understanding of the connotation based on big data, based on the transformation of services concept, and using correlation method Mapreduce and data mining of large data accurate to push information to readers to improve the quality of services. In order to achieve the promotion of Readers information service capabilities, thus promoting the development of China's Librarianship.

First proposed the big data era has arrived who is McKinsey: data has penetrated into every area of today's industries and business functions, has become an important factor of production. People for mining and use of vast amounts of data, heralding a new wave of productivity growth and consumption wave arrivals surplus [2]. Big data and massive data are different, not only refers to large amount of data, but a huge amount of information refers to the amount of data involved in large scale to the human brain can not even through mainstream software tools to achieve within a reasonable time capture, manage, process, and organize to become more actively aim to help business decision-making information. Big data feature can be planned as four "V", they are: Volume, Variety, Value, Velocity.And they have four-level characteristics: First, the data volume is huge; second, the data type are many kinds. Third, low value density, high commercial value. Fourth, fast processing speed [3].

Occupies a very important place in society, not only because it is the hub distribution center and disseminated literature and information resources of literature information resources, but also it can

guide and help the reader establish a correct outlook on life and values, and lay basicscientific theory to establish a political direction for the construction of socialism with Chinese characteristics and strive [4]. With the rapid development of information technology , resource content of university involved more and more areas, involving multiple carriers in the form of resources, thus the university will face to handle massive resources. The traditional service mechanism is based on a single form online retrieval , and information retrieval methods often through a combination of keywords, matching the corresponding document in the database andreturning it to the reader a way to do it, but the local match refers the user to retrieve the contents of a partial match, rather than an exact match, it often can not meet the real needs of users. Although at present there are for full text content for retrieval of full-text retrieval technology, they still rely to user's query request, no in-depth mining the user's characteristicsand retrieval of habits, not true to the user to provide the best service. Therefore, recommendation services receives a wider range of attention,it can tap users information, books information and users operation logs and other hidden correlation between data mining, To carry out active retrieval without the user submit the query content, active recommendationinformation to the user.

2. User Mode

2.1. MapReduce Computational Framework

The typically representative recommendation service is recommendation service in . recommendation is defined according to the user's previous preferences and browsing behavior, through the establishment of users model to match the user-related information to the user and the information to recommend post-treatment[5]. This technique has been widely used in electronic commerce, mainly due to expanding the number of species and commodities rapid growth of e-commerce model, resulting in users will spend a lot of time to filter product information in order to find their own goals merchandise. In this process, the user can do lots of wasted efforts, and even can not find a suitable target, causing the target user's losses. Digital emergence of recommendation in this regard is also the reason, hoping to save the reader's time by the initiative of recommendation information to the reader, so that readers can rapidly detect the target group of interest. recommendation system is a method of using a variety of data mining knowledge database to dig out the knowledge of applied technology, its goal is to provide real-time, proactive and valuable information to the user. In addition, The system can also be recommendation to the satisfaction of the user feedback, the statistics of users, in orderto verify the effect of recommendation. In totally, we can use specific data mining based on bigdata preprocessing, and some pre-excavation of the to access the data, mining the usersmode.

3. The Information Recommendation Strategy Based on Big Data

3.1. Customer Hierarchical Design

Hierarchical design is mainly in order to distinguish between different users to provide service, the advanced users enjoy more services.

According to the characteristics of users, which can be divided into ordinary users domesticuser and advanced users advanceduser. Ordinary users and advanced users need to consider many factors including the spend of access to the is duration,visit the freshness of the as new,and frequency of access to the as Frequent.

Where count is the number of times a user access to the , durationi for i time users spent a lot of time to visit the .Set up a user first accesses to the Business date as userstart, the current date is usercurrency.

Freshness new is a relative value, reflecting the freshness of user access time to the , is described as:

$$new = \frac{userend \quad userstart}{usercurrency \quad userstart}$$

calculate the current user's freshness of new, the spend time of visit to the as duration, access to

libraries frequency frequent;

for the value of each calculated using fuzzy function converts to the [0,1] range; frequent

weight maximum, set α;

duration greater weight, are set β;

new minimum weights, are set γ;

user threshold value is set θ;

If (*frequent + *duration + * new>)

The user is advanced users: advanceduser;

Else The user is a normal users:

domesticuser;

3.2. The Users Recommend Strategy

Using data mining technology is different from the conventional information finding, its main task is to find the hidden relationship between large amounts of data from a large number of structures, vague, many types of fuzzy data mining knowledge, and to build a model for them. Recommendation strategies based on big data are the following four ways:

recommendation based on association rules: Its main idea is to tap the implicit between things in the database project association. It has been associated by the rules. The advantage is that rules can be calculated offline, to enhance the real-time corresponding capacity, and it can find links between different types of projects, its shortcomings are: generation rule needs to repeatedly scan the entire database, the operation efficiency is not high, because the algorithm offline execution, it does not update the entire rule base in real time; it is good at finding links between different types of projects, but can not find a good contact between similar types of projects, the new projects recommendation model is very difficult to detect[6].

Classification algorithms: it's learning through the training data, the data is divided into different categories, and each category has its own concept description, classification process, while establishing a classification rule, and then, using the same database test samples to verify the validity of the rule model[7]. A valid rule model can better serve the estimated classification of new data. In the , the classification algorithm can borrow the log file reader calculated to discover which books readers to compare preferences. Tree is commonly used classification algorithm. The process is based on the greedy algorithm achievements, which are used for commercial purposes.

Clustering algorithm: It instance all the data into a number of individual groups met, these different groups are called clusters. Clustering same individual has more consistent properties, individual similarity between different clusters of small[8]. K-means method is the most typical representative, it does not require training samples without knowing the properties characteristic of each group. Its algorithm is based on iterative implementation process, the disadvantage is that the object data distribution rules for noise-sensitive and can only clustering. As a representative simple clustering algorithm efficiency of unsupervised learning, it has been widely used in various fields of life. Recently, with the rapid development of online text, online text clustering has become an important demand for network readers.

Time series: Data will be processed by time order, and find a higher probability of occurrence of the sequence mode. Its main idea is to analyze and forecast based on existing data. In the , which is mainly used to analyze the reader to borrow time, the frequency of different periods by borrowing to know the rational allocation of human resources and borrow time, thereby improving the quality of services.

Based on each recommendation algorithm to build the recommendation system, we can be rational technology applied to of data mining, the recommendation. Due to the ordinary users of the are less collected of their personal information, According to the ordinary users, proposes the Mapreduce and user mode recommendation method based on Clustering as follows: This method uses Mapreduce based on big data technology, Firstly, the similarity of the user model clustering to generate user clustering models; then current user access path and user clustering pattern matching; Finally, according to the set threshold to generate the recommendation set, the recommendation set to the

user. The method uses similar interested users to obtain knowledge recommendation sets, recommendation speed is very fast, suitable for ordinary users.

While advanced users is the key of service, object comparison with a higher value, and the user is also very important. In order to attract and retain high-level users, the can make recommendation method based on Mapreduce and recommendation based on association rules. This method uses Mapreduce technology based on big data, Firstly, mining frequent access paths from the user mode, and then dig out the use of frequent access path to match the user's current association rules; Finally, according to the association rules and the recommendation degree factor to determine the recommendation set, the recommendation set is recommendation to the user. The proposed method is highly accurate and suitable for the advanced users of .

On the basis of the user layer and the service method of the under the big data, the strategy of information recommendation under big data is put forward. The information recommendation strategy is described as follows:

Calculate the current user when accessing librarian duration, freshness new, visit the frequently frequent;

for each value, use fuzzy function switch to range [0,1];

frequent weight maximum, set α;

duration greater weight, are set β;

new minimum weights, are set γ;

User threshold value is set θ;

If (*frequent + *duration + * new>)

The user for advanced users: advanceduser;

Else

This user is a normal users: domesticuser;

If (current user as an ordinary user domesticus)

Make recommendations based on the use of recommendation Mapreduce user mode clustering methods;

Else To recommend by recommendation method of recommendation based on Mapreduce and association rules mining.

4. Experiments

Test environment for an HP server. The experimental data from the log file of the , the total data amount was a month for access to records, data preprocessing. Choose 20 users to evaluate the accuracy of the page to evaluate the analysis, each user to carry out two tests. The first time wasbased on the Mapreduce and user mode clustering of recommendation method to test, the second times of recommendation method based on Mapreduce and association rules mining to test. 20 customers through two tests to evaluate the accuracy of recommendation pages. Average accuracy obtained were shown below.

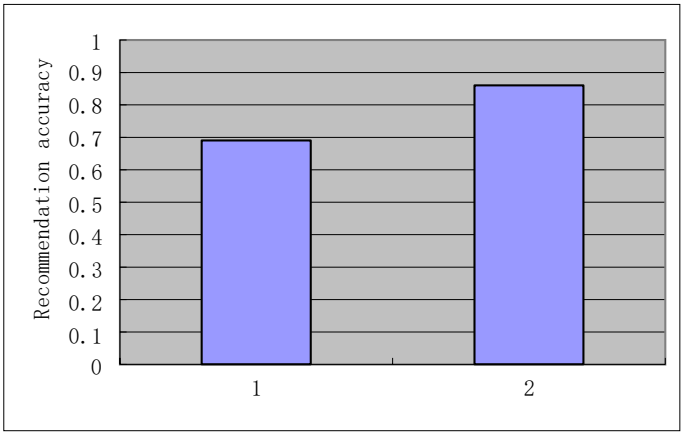


Fig 1. The recommendation accuracy of two recommendation methods

The accuracy of two recommendation algorithm are 0.67 and 0.86. High accuracy was based on MapReduce and association rules mining of recommendation method, the method for advanced users, need a more accurate recommendation. Low accuracy was based on MapReduce and user model clustering of recommendation method, the method for the general customer, when recommendation considers more of the efficiency of the The experimental results show that both of the recommendation method is suitable for the user groups, based on the user recommend a layered strategy fully consider the characteristics of two different customers. Better services for different users.

5. Conclusion

The recommendation can hide information on the depth of excavation, to obtain users information that users can not obtain by the information retrieval results. Because intelligent and humane service, making it application area more and more widely, recommend the use of technology has become even more and more reasonable. Paper presents a of recommendation service based on user-tiered strategy. The experimental results show that the proposed strategy is effective.

References

[1] Han J W, Micheline K. Concept and technology of data mining[M]. Beijing: Machinery Industry Press, 2013.

[2] Chen ZB, Han H, Wang JX. Data Warehouse and Date Mining [M]. Beijing:Tsinghua University Press,2009.

[3] Zeng,Y,Yin S,Liu J,Zhang M. Research of improved FP-growth algorithm in association rules mining [J]. Scientific Programming, 2015, 46(10),281-285.

[4] Yan Yun, YX Luo. Improvement of FP-Tree algorithm [J]. Computer engineering and design, 2010,31(07):1506-1510.

[5] Rathi,Sheetal,Dhote,C.A. Parallel implementation of FP growth algorithm on XML data usingmultiple GPU[J]. Advances in Intelligent Systems and Computing, 2015.8, 339,581-589.

[6] Zeng C, Xing CX, Zhou LZ. A survey of personalization[J]. Jounal of Software, 2002, 13 (10): 1952~1961 (in Chinese with English abstract).

[7] DEAN J,GHEMAWAT S. MapReduce:simplified data processing on large clusters[J]. Communications Of The ACM, 2008,51(1):107-113.

[8] SH Li, ZW Lv, DY Che. Maximum frequent itemset mining algorithm based on ordered FP-tree[J]. Northeast Normal University newspaper (NATURAL SCIENCE), 2016,48(2):65-69.