Academic Journal of Engineering and Technology Science
ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

An Overview of the Application of Data Mining Technology in E-commerce

Yongsai Yan, Qiwei Liu

School of Software, Jiangxi Normal University, Nanchang, 330000, China

ABSTRACT. *Data mining technology is an effective data analysis and processing technology. It analyzes and processes sales data according to corresponding association rules, classification, clustering, and prediction techniques, and discovers the hidden knowledge in the data, which in turn can be a marketing strategy. Decision making and product development provide decision-making and have a wide range of applications in e-commerce. This paper first expounds the data mining technology and e-commerce concept, then analyzes and summarizes the research results of domestic and foreign scholars, the application status and development of data mining in e-commerce, and finally the application of data mining technology in optimizing e-commerce website design. And the application in business customer relationship management is fully discussed and researched to explore the convergence of e-commerce development and data mining, and comprehensively promote technological innovation and application.*

KEYWORDS: *e-commerce, data mining, technology research, clustering technology*

1. Introduction

With the advent of the era of big data, traditional data mining algorithms have become less and less satisfactory. We need to develop more efficient data mining algorithms and tools to handle large amounts of data of different attributes, types and dimensions to support the right management and decision making. The rapid development of e-commerce has led to many problems. A common problem faced by all companies today is that although the e-commerce system collects a considerable amount of data, there is very little real value that can be obtained. How to effectively organize and utilize this information and obtain information that is conducive to business operations and enhance competitiveness from massive data is the focus of the enterprise.

Data mining technology with statistics as the main background has played a huge role. In the process of enterprise operation, the database system is gradually being widely applied to daily information management, which is to effectively organize and utilize the massive information in the database through relevant technical means

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

to help relevant data owners find valuable information and knowledge. Guide the development planning of the enterprise and improve the operational efficiency of the enterprise. Data mining technology can help e-commerce operators solve data analysis and mining problems. Data mining technology is an effective data analysis and processing technology. It analyzes and processes sales data according to corresponding association rules, classification, clustering, and prediction techniques, and discovers the hidden knowledge in the data, which in turn can be a marketing strategy. Decision making and product development provide decision-making and have a wide range of applications in e-commerce.

Data mining technology can discover the rules from the data with the above characteristics, track the browsing behavior of customers on the Web and conduct pattern analysis, shorten the distance between enterprises and customers, enable enterprises to better understand customer needs, and develop e-commerce with lock-up activity. Data mining for e-commerce is a typical application of Web data mining. At present, the convenience and fast transaction speed brought by the business activities through the Web has become the key driving force for the rapid development of e-commerce. On the other hand, various types of business activities involving clients are undergoing tremendous innovation. E-commerce websites generate millions of online transactions every day. How to analyze and mine a large amount of relevant data related to transactions is import for fully understanding customer preferences and purchase patterns, assisting business management decisions, and designing products that meet different customer groups. Promotional models and personalized website services increase corporate competitiveness, which is the research field of Web data mining.

2. Data Mining Technology Overview

2.1 Introduction to Data Mining Technology

From the perspective of decision control, data mining can be understood as a decision support process, which can automatically analyze all the data involved in business operations, make inductive reasoning, and explore potential patterns and predict customers. Behavior, helping enterprise decision makers adjust market strategies, is the main algorithm used in data mining systems is the application of knowledge discovery technology in artificial intelligence. At present, data mining research and development shows that data mining needs to cover a variety of application tasks, from data preprocessing to association rule discovery, cluster analysis, data classification, deviation checking, sequence pattern analysis and other specific tasks. Therefore, data mining applications are a challenging application area that combines multiple disciplines and technologies.

Data mining technology is considered to have exciting research prospects because it can be used in a wide range of commercial applications. Such as to support the decision-making of enterprises, the formulation of market strategies, and so on. Faced with the influx of large amounts of data, enterprises have greatly

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

demanded the application of data mining, which will enable this technology to be rapidly developed and improved. In foreign countries, large enterprises such as large-scale commercial enterprises, financial industry, insurance industry and civil aviation have begun to apply, but the country is still in the initial stage of theoretical discussion and application.

Data mining has two tasks: (1) database understanding of VL: transforming the database into a more concise model that can be understood by the computer in the representation, and then using this model to solve new problems; (2) human database understanding: Simplify the data as needed and translate it into natural representations (such as mathematical formulas, natural language and diagrams, etc.), and discover the rules that are implicit in large amounts of data and make them understandable. Through data mining, rules can be directly exported from the instance data and be used to construct the knowledge base. And the existing rules can be verified in the database, so it is also essential to check and update the knowledge base. In control theory, the expert system has been studied for a long time, and quite a lot of research results have been discovered and applied to the actual system.

2.2 Data Mining Technology Features

Data mining is used to discover implicit, meaningful knowledge from a database. Today's data mining capabilities and the types of knowledge they can discover are as follows:

2.2.1 Discovering general knowledge through conceptual description

There are many methods and implementation techniques for concept description, such as data cubes, attribute-oriented induction, and so on. There are other aliases for data cubes, such as "multidimensional database", "implementation view", "OLAP", and so on. The basic idea of the method is to implement calculations of some commonly used in costly aggregate functions, such as counting, summing, averaging, maxima, etc., and store these result implementation views in a multidimensional database. Another generalized knowledge discovery method is the attribute-oriented induction method proposed by Simon Fraser University in Canada. This method expresses data mining queries in a SQL-like language, collects relevant data sets in the database, and then applies a series of data promotion techniques on the relevant data sets for data promotion, including attribute deletion, concept tree promotion, attribute threshold control, and counting. And there are other aggregate function propagation.

2.2.2 Correlation knowledge is discovered through correlation analysis

Data correlation is an important discoverable knowledge in database. If there is some regularity in the relationship between the values of two or more variables, it is

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

called correlation. Correlation can be divided into simple correlation, temporal correlation and causal correlation.

2.2.3 Discover classification knowledge through classification and clustering methods

Classification knowledge is the characteristic knowledge of the common nature of the same kind of things and the difference characteristic knowledge between different things. The most typical classification method is based on the decision tree classification method. The clustering method is to group data objects into multiple classes or clusters, and the objects in the same cluster have higher similarity, and the objects in different clusters have larger differences. Clustering differs from classification in that the classes it is to classify are unknown. The similarity is calculated based on the attribute value of the description object. Cluster analysis is widely used in pattern recognition, data analysis, image processing and market research.

2.2.4 Obtain predictive knowledge through prediction methods

Predictive knowledge refers to the estimation of future data from historical and current data based on time-series data. It can also be considered as related knowledge with time as a key attribute. At present, time series prediction methods include classical statistical methods, neural networks, and machine learning. At the same time, a re-training method based on statistics and accuracy-based is proposed. When the existing prediction model is no longer applicable to the current data, the model is retrained, new weight parameters are obtained, and a new model is established. There are also many systems that use the computational advantages of parallel algorithms for time series prediction.

2.3 Key technologies for data mining

2.3.1 Neural network

The neural network is a nonlinear prediction model modeled on the physiological neural network structure, and learns pattern recognition. Because it provides a relatively simple and effective method for solving complex problems, it has received more and more attention in recent years. Neural networks are often used in two types of problems: classification and regression, based on a self-learning data model. It analyzes large amounts of complex data and performs extremely complex pattern extraction and trend analysis on the human brain or other computers. Neural networks are well suited for both nonlinear and noisy data, so they are widely used in the analysis and modeling of market databases.

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

2.3.2 Decision tree

A decision tree is a predictive model of a tree structure in which non-terminal nodes of a tree represent attributes and leaf nodes represent different categories to which they belong. The branches of the tree are established according to the different values of the data in the training data set to form a decision tree, which is repeatedly pruned and converted into rules. The decision tree is the process of classifying data through a series of rules. A typical example is the CART regression decision tree approach.

2.3.3 Statistical analysis

Establish two relationships between the database or other datasets, that is, functional equations can be used to represent functional relationships that determine relationships and correlations that cannot be expressed by functional formulas but have related relationships. Regression analysis can be used for their analysis, correlation analysis, principal component analysis, etc. In the actual process, these technologies are usually not used alone. People use a variety of data mining techniques to achieve the best results. For e-commerce, we will introduce it in the next chapter and apply these data mining techniques to e-commerce systems.

2.3.4 Classification analysis

Classification is to find a conceptual description of a category, which represents the overall characteristics of such data, that is, the meaning of the category, generally expressed in rules or decision tree patterns. The connotation description of a class is divided into a characteristic description and a distinctive description. The characteristic description is a description of the common features of the objects in the class; the distinctive description is a description of the difference between two or more classes.

2.3.5 Cluster analysis

The data in the database can be divided into different classes according to certain rules. In the same category, the distance between individuals is small, while the distance between individuals in different categories is too large. The difference between cluster analysis and classification analysis is that the data in the database before the cluster analysis does not contain any category markers, and the data tuples with common trends and patterns are aggregated into one class, so that the tuples in the class have the highest degree of similarity. The biggest difference lies in practical applications. Based on the data of existing customers, cluster analysis can be used to divide the market into several sub-divided markets according to the similarity of customers' consumption patterns, and to develop targeted market-field strategies to improve sales performance.

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

3. E-commerce overview

3.1 Definition of e-commerce

E-commerce is called electronic transaction (E.commerce) in a narrow sense. It mainly refers to commercial trade activities conducted on the Internet by means of Web communication; in the broad sense, it refers to all business activities using the Web including electronic transactions, also known as E-business, such as market analysis, customer management, resource allocation, corporate decision-making, etc. In summary, e-commerce refers to various business activities, trading activities, financial activities and related comprehensive service activities carried out through the information network in a wide range of commercial trade activities around the world.

3.2 Classification of e-commerce

B2B (Business to Business): refers to various business activities between the enterprise and the enterprise using the network. Traditional business transactions often consume a lot of resources and time, and the cost of products is occupied by sales, distribution, and procurement. Through B2B trading, both buyers and sellers can complete the entire business process online. B2B reduces transactional and administrative costs and reduces operating costs by reducing transactional transactions between companies.

B2C (Business to Consumer): refers to the e-commerce activities between enterprises and consumers, mainly business activities such as online ordering. Business-to-consumer transactions are largely retail, and there are already thousands of forces on the Internet. The online shopping website provides a variety of consumer products, such as the famous Amazon website (http://www.amazon.com) and the domestic Dangdang website (http://www.dangdang.com). In the long run, the B2C website can enable companies to increase communication with customers, provide customers with more choices, and provide more personalized services that are not possible with traditional business methods.

B2G (Business to Government): refers to the affairs of a company involving a government agency, including all transactions between a business and a government agency, such as corporate online customs declaration, online tax filing, online application for license or business license Wang, online property rights transactions, government online bidding, online procurement and other activities.

C2G (Consumer to Government): refers to the affairs of individuals involving government agencies, mainly including personal identification, tax filing, tax collection and other government-to-person transaction processing through the network.

4. Data mining techniques in e-commerce application condition and

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

development

Data mining technology has been closely related to Knowledge Discovery in Database (KDD) since its inception. In 1989, Piatetsky. Shapiro and Frawley's essays (Knowleadge Discovery in Database) brought together some of the early papers related to data mining. After that, KDD's research focus shifted from discovery methods to system applications, with a focus on multiple strategies and techniques., integration and inter-disciplinary penetration. The papers published in 1996, "Advances in Knowledge Discovery and Data Mining" (edited by Fayyad, Piatetsky. Shapiro, and Smyth et al.), have begun to show some in-depth research on data mining technology. Data mining and knowledge discovery technology have begun to play its part in the advantages and effects of processing massive data.

At present, the application research of data mining technology in e-commerce website design focuses on optimizing website link structure, page real-time recommendation and personalized site design. Since the user's access behavior on the website is saved by the web server in the form of a log. Mining the web log to discover the user access pattern to optimize the website design naturally becomes the mainstream in this field. Mobasher et al. proposed a page recommendation algorithm based on association rules and cluster analysis [20], which can be used to build personalized recommendation websites. In the literature [20], Bose et al. first proposed a method to optimize the site structure by adding hotlink. On this basis, Fuhrmann et al. used this method in the literature [21] to optimize the site link structure, that is, the path distance between the node with higher weight and the root node is minimized by the addition of a limited hotlink. In [10], Cooley et al. define the interest degree of frequent itemsets based on user access frequent routes, and use the path with higher interest as a reference for improving the structure of the site. In [11], Edmond et al. introduced the interest degree index into the session-based sequence pattern, but it was not used to modify the website structure, but to propose an algorithm that can reduce the hyperlinks searching for these indicators. Many scholars in China have also proposed an optimization method for improving Web sites by introducing clustering analysis, association rules, sequence patterns and other mining algorithms in Web log mining.

There is a close relationship between CRM and e-commerce. In recent years, many CRM systems have been used in many European and American countries. Many domestic companies have begun to pay attention to modern marketing concepts and business operations to the use of CRM, and also make data mining in CRM at home and abroad. The research has a very wide range of values. The research on data mining algorithms and models for CRM customer identification, customer segmentation and customer retention has greatly promoted the application of data mining in the field. Scholars in the international academic community mainly use the relevant algorithms of data mining technology to establish a market-based customer segmentation model. For example, Morwitz studied the use of CART algorithm, K-means clustering method and discriminant analysis method to segment the customer market in the literature [17], and analyzed the efficiency and effectiveness of the three algorithms. In the literature [3], Jaesoo Kim et al.

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

studied the application of neural network algorithm in tourism customer segmentation and explored the advantages of neural network application in customer segmentation. As far as the research on the application of data mining in customer relationship management is concerned, it is obviously lagging behind compared with other countries, but it has made a lot of progress in the research of application algorithms and application fields in our country.

The combination of data warehouse solution and OLAP technology is also a research hotspot in the field of data mining. The data processed by extraction, conversion, cleaning, etc. can be better identified by data mining algorithms, and OLAP can target one of the data warehouses. The subject conducts online data access, processing, and analysis, accessing information from multiple perspectives quickly and consistently, meet the specific query and report needs of decision makers in a multidimensional environment. Many scholars at home and abroad have developed data warehouses for e-commerce data mining and the use of OLAP to research the results of visual mining. Although there are no mature products at home and abroad that combine e-commerce and OLAP, the emergence of data mining tools provides solutions for data mining problems in the e-commerce field.

5. Application of data mining technology in e-commerce

5.1 Web usage mining process in e-commerce website design optimization

The mining process of Web usage pattern mining is the same as the Web mining process. It is simply divided into four phases: data collection, data preprocessing, data mining, and analysis of the mined patterns, as shown in the figure.

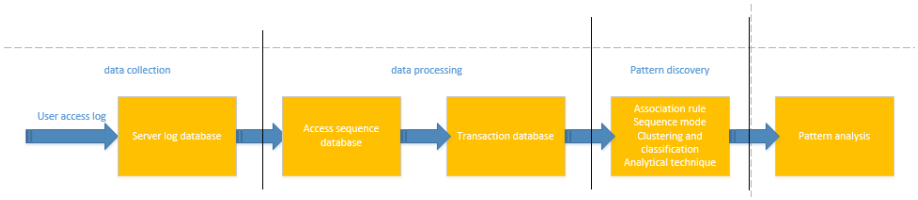

Figure. 1 WEB usage pattern mining process

The user access information is recorded in the server, which is the main data source for Web data mining: the log data is preprocessed to form a transaction database, which is the basis of data mining. Then, it is supposed to use the mining rules, such as association rules, sequence pattern, classification and clustering, to carry out data mining. Finally, the paper analyzes the results of the mining and optimizes the website design with mining results.

The commonly used techniques in Web usage pattern mining include the unique path analysis technology used by the Web, the association rules, sequence patterns,

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

classification clustering techniques commonly used in the data mining field, and Path Analysis Technique. The most commonly used material in Web-based record mining is graphs, which represent the relationships between web pages defined on a website. A page is defined as a node, and a hyperlink between pages is defined as an edge.

Web-based use of record mining is to find the most frequent access path from the graph. This allows path analysis to determine the most frequent access paths on the site.

Sequence mode: The visitor browses the website or uses the website to provide services. Sometimes the sequence relationship is applied. By applying the sequence pattern, the next action can be predicted according to the current visit behavior of the visitor, and the corresponding website content is provided according to the predicted result. Services refer to adding a hyperlink or providing a recommendation link to a web page, or placing a different banner on a page for a specific user group to increase the click rate of the advertisement, and the like.

Discovery of association rules: Some web pages or services are always used or browsed by visitors, but there is not necessarily a hyperlink relationship between them, so association rules can be used by operators as a reference for improving site organization and structure.

5.2 Application of Web Usage Data Mining Technology in Optimizing Website Design

The functional modules designed according to the Web mining process can solve many practical problems in the website design optimization and operation process, such as: optimizing the site link structure, improving the site content structure, clustering access users and so on.

Optimize website page design. The content settings of the website page directly affect the access efficiency of the website. Also, the content of the site's visitors will change at any time. The mining algorithm of the above mining tool mines the statistical information of the user’s accessing the page in the log file of the network, and finds the mode used by the user, which can provide good suggestions for improving the content setting of the webpage.

In addition, establish an association matrix with the URL of the W-cursor as the row, the USERID as the column, and the value of element is the incidence matrix of the number of visits by the user in a period of time:

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

$$M_{m\times n}=\begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1j} & \cdots & l_{1n} \\ l_{21} & l_{22} & \cdots & l_{2j} & \cdots & l_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{i1} & l_{i2} & \cdots & l_{ij} & \cdots & l_{in} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ l_{m1} & l_{m2} & \cdots & l_{mj} & \cdots & l_{mn} \end{bmatrix}$$

Each row vector L[*,j] represents all client accesses to the URL "*"; each column vector L[*,j] represents the client "*" access to all URLs in the commerce site. Therefore, it can be considered that the column vector is a personalized sub-picture of the customer accessing the site, and the customer having the similar access sub-picture is the customer group having the same browsing mode. When a new user visits a website, the website can dynamically adjust the content and structure of the website for the customer, create a customized page that highlights the product, and make the connection between the related files accessed by the customer relatively straightforward, so that the customer can easily access the website. Visited page.

5.3 Analysis of customer data characteristics in e-commerce system

Data mining technology can be applied in e-commerce customer relationship management. The e-commerce environment established to solve the problems of T (delivery), Q (quality), C (cost), S (service), and E (environmental protection) of enterprises covers the entire process of production and operation activities of enterprises. It integrates personnel, technology, management and enterprise logistics, capital flow, and information flow, and is an organic whole. On the basis of examining the management issues of enterprises, the author believes that the comprehensive information of enterprise management has the following characteristics:

(1) The amount of data is large and wide. Comprehensive information on enterprise management includes financial management, employee management, energy management, material reserve management, order management, production schedule management and cost management, and the source of this information is multifaceted.

(2) The data is in various forms. There are traditional files, database data, Internet/intranet data, and multimedia data such as sound and images. The same type of data (such as database data) may also be heterogeneous with a completely different storage format.

(3) For different businesses. Different information is provided for different services. Usually, different services have different emphasis on information needs.

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

They need information that is directly related to their management and a large amount of auxiliary information.

5.4 Data Mining Application Process in Customer Relationship Management

The application of data mining in customer relationship management is to extract the business data, marketing data and customer data in the enterprise data warehouse as modeling samples, use various data mining methods to mine the data, analyze and build the mining results. Models continue to optimize the model to guide business decision-making and market planning. The process of data mining applied in customer relationship management is as follows:

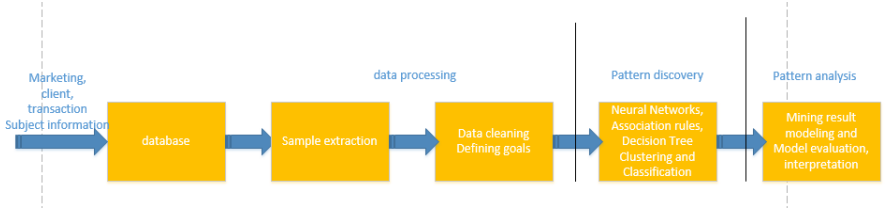

Figure. 2 Customer data mining process

The data related to customer relationship management in the source data collected by the data collection process includes server log data, customer registration information, transaction data information, etc., wherein the combination of customer information and server log data can better understand customer behavior. The combination of customer information and transaction information can analyze the response of different customers to the product.

6. Conclusion

First of all, this paper makes a detailed exposition in the introduction about the function and key technology of data mining technology, and further explains the classification of e-commerce so as to better explain the connection and difference between the two. At the same time, when researching data mining technology in e-commerce application, it is divided into two aspects: the application of data mining technology in optimizing e-commerce website design and the application of data mining technology in e-commerce customer relationship management; This paper further elaborated Web use mining process in e-commerce website design optimization, Web use data mining technology in optimizing website design, customer data characteristics analysis in e-commerce system, data mining application process in customer relationship management.

Secondly, due to the time limit, ability limit and weakness in many aspects, there are still many areas worth learning and upgrading; some data mining algorithms are also used, although it can be found in the research of many scholars. When solving

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

some of these problems, the same algorithm can be used. But since data mining in e-commerce is an extremely complicated process, there are many unresolved problems, and a single data mining algorithm can not complete diversified data well. To mine the demand, only by combining the effective multiple algorithms and integrating them into the e-commerce system can we exploit the advantages and functions of data mining to process large amounts of data.

References

[1] Zhu Wenxiang. Research and application of data mining technology in B2C e-commerce [D]. Hefei: University of Science and Technology of China, 2011.
[2] Ren Xin. Web data mining and its application in e-commerce [D]. Guiyang: Guizhou University, 2008.
[3] Shen Hongchao. Application of Data Mining Technology in E-commerce [D]. Nanning: Jiangnan University, 2009.
[4] Feng Yongping. Application of data mining technology in e-commerce [D]. Chengdu: University of Electronic Science and Technology, 2012.
[5] Zhu Wenxiang. Research and application of data mining technology in B2C e-commerce [D]. Hefei: University of Science and Technology of China, 2011.
[6] Wang Zhongzhuang, Deng Lundan, Shi Wenbing. Application of data mining technology in e-commerce recommendation system [J]. Microelectronics and Computing, 2007, 24(4): 197-199.
[7] Liang Xiexiong, Lei Yihuan, Cao Changxiu. Research Progress in Modern Mining Technology [J]. Journal of Chongqing University, 2004, 27(3): 21-27.
[8] Yang Jinlu. Application Research of Web Data Mining Technology in E-commerce [J]. Electronic Technology and Software Engineering, 2018(2): 118-118.
[9] Huang Zhiheng, Gong Qin Application Research of Data Mining Technology in E-commerce[J].Information Technology,2017(34):115-116.
[10] Cheng Junfeng Application Research of Data Mining Technology in E-commerce[J]. Journal of Anyang Teachers College, 2015(1): 40-43.
[11] Chen Li, Jiao Licheng. Research Status and Latest Development of Internet/Web Data Mining[J].Journalof Xidian University(Natural Science Edition)2001,28(1):114-119.
[12] FAYYAD U, PIATESKY-SHAPIRO G, SMYTH P. The KDD Process for Extracting Useful Knowledge Form Volumes of Data[J].1996,39(11):27-35.
[13] FAYYAD U. Mining Databases:Towards Algorithm for Knowledge Discovery[J]. 1998,21(01):39-48.
[14] PARK J, CHEN M, YU P. An Effective Hash Based Algorithm for Mining Association Rules[J]. 1997,9(05):813-825.
[15] ZAIANE O R. Resource and Knowledge Discovery from the Internet and Multimedia Repositories[D]. 1999.
[16] DUNJA M. Text-Learning and Intelligent Agents[R]. Slovenia, Jozef: Stefan Institute,1998.
[17] BURKE R, HAMMOND K, KULYUKIN V. Question Answering from Frequently Asked Question Files[J]. 1997,18(02):57-66.

ISSN 2616-5767 Vol. 2, Issue 2: 76-88, DOI: 10.25236/AJETS.020044

[18] MAES P. Agents That Reduce Work and Information Overload[J]. 1994, 37(07):30-40.
[19] MITCHELL T, CARUANA R, FREITAG D. Experience with a Learning Personal Assistant[J]. 1994,37(07):81-91.
[20]HEDBERGS.Agentsforsale:firstwaveofintelligentagentsgocommercial[J].1996,6 (6):16-19.
[21] BALABANOVICM, SHOHAMY. Fab: Content Based, Collaborative Recommendation[J]. 1997, 40(03):60-70.
[22] Charles E Goldfarb, Paul Prescod, devlopment and application of compute, the implement of internet commerce,2000.3.