Meta-analysis of network information security and Web data mining techniques

Dongling Wu[1, a], Shaolong Shan[1,b]

[1]Tangshan Vocational & Technical College, Tangshan, China

[a]176288793@qq.com, [b]29463989@qq.com

Keywords: Network Information security; Data mining; Scarcity defence model

Abstract. With the development of network economy, network and information security has become an important factor in the further development of the network economy. Web data mining technology is a key technology to enhance the performance of network information security. It can effectively improve network and information security. Based on this, we study Meta-analysis of network information security and Web data mining techniques deeply.

Introduction

In recent years, with the rapid spread of the Internet, the network is becoming a very important and indispensable means for users to complete the relevant business. The network economy is born based on commerce, either in foreign or domestic, which have been leaps and bounds. On the other hand, the current network security status that the network economy faces is not optimistic. [1] Network and information system and its inherent disadvantages, vulnerability and threats, making network security has become not only an important part of country and national defense security, but also an important bottleneck restricting the further development of the network economy. Network and information security problems, both theoretically and technically are not completely resolved, therefore, we can not be to combine the network information security technology with other technologies and based on existing historical data, to enhance network and information security targeted prevention, timeliness and effectiveness? The answer is yes. This technology is the Web data mining technology which enhances the performance of network information security offers the possibility and feasibility.

Web data mining

Data mining is from a large, incomplete, noisy, fuzzy, random data to extract implicit in which people are not known in advance, but is potentially useful information and knowledge[2].However, most of the traditional data mining methods only for homogeneous, are omorphic to analyze the data, which for a large number of heterogeneous text information on the Internet, log information, hyperlinks, etc. is not applicable.

To solve this problem, some people take some measures to combine traditional data mining technology with Web to produce a new mining technology --- Web data mining. Web data mining is a technology. It can discover and extract some potentially useful patterns of interest and hidden information from Web documents and Web activities. It mining useful information from the Web as the goal, and it based on data mining, document mining, multimedia mining and use of computer networks, databases and data warehouses, artificial intelligence, information retrieval, visualization, natural language understanding technology, which will pass data mining combined with Web[3].

The basic process of Web data mining is shown as Figure 1.

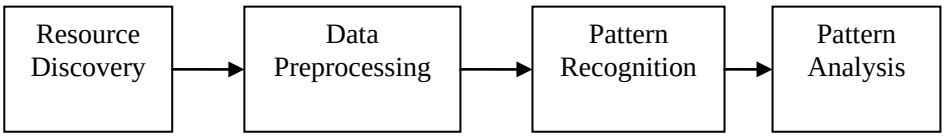

Figure 1 The basic process of data mining

Among them, the resource discovery refers to the process of acquiring and return text resources from Web. The object of their treatment include static Web pages, Web databases, Web architecture, user records and other information.

Meta-analysis

Web data mining model based on network information security. Network information security model based on Web Data Mining is shown as Figure 2:

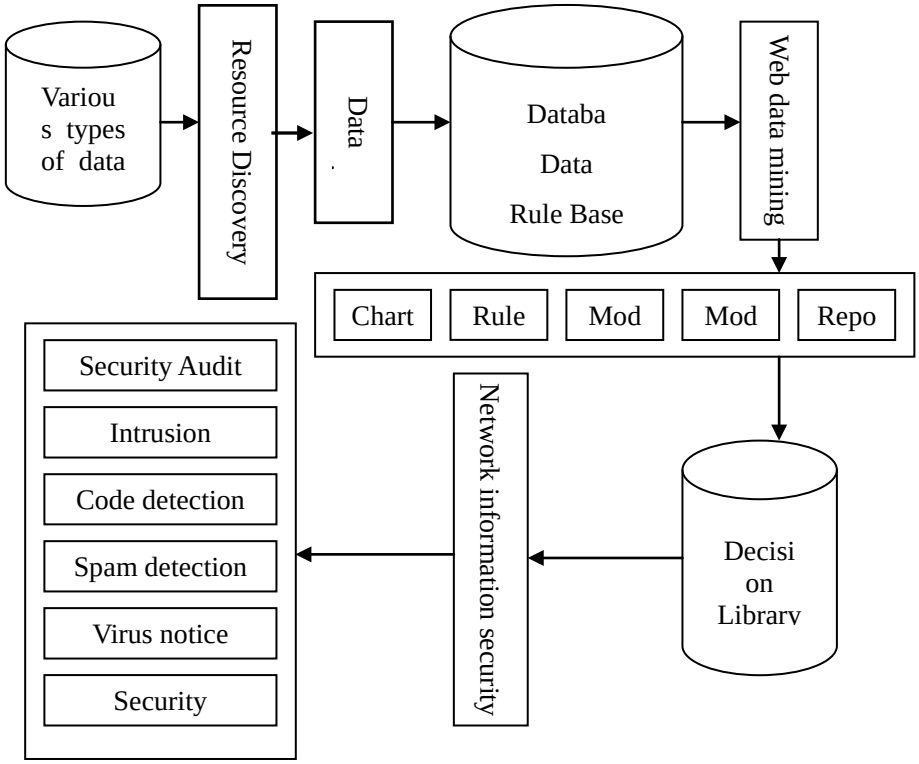

Figure 2 Network information security model based

As a comprehensive analysis tool of network IT security, mainly in three areas[4]. First is Filter. Its role is to extract some of the data in the database, and then use ambiguous methods of analysis than the right, so that this information consistency. Second is mining synthesizer. This is an engine driven digging means. Its role is based on mining claims, mining system to select the appropriate method to carry out excavation in the algorithm library. Third is method selection expert system and knowledge base.

These are an important part of Web data mining, data mining system according to the specific requirements of the user to choose the most effective mining algorithm, along with technical knowledge but also through the development of Web content and constantly updated rules to improve intelligent systems.

Data Mining Prevention Model. The first one is security Audit. Security audit system is targeted for some security-related data and records generated on the network for analysis and statistical process. Security audit is the process of analyzing incidents involving safety and recorded it for

security events are: user operating system, user network activity, system applications and services, network behavior. Web technology can help network security audits.

The main principle is to dig in the normal data to normal network traffic patterns, these communication patterns and after a number of attacks linked to the associated rule base analysis, the detection system through its post-analysis will detect the number of potential vulnerabilities, found that the program is a security problem, and then take the appropriate action to resolve. Web data mining through some of the technical and safety audit system integration can turn on the HF firewall, IDS intrusion systems to protect information, the timely detection of the security status of the network, for staff to provide data for timely information and systems the current operating status.

The second one is Intrusion Detection. Some of the behavior of the user through the information gathering and analysis, if abnormal behavior is abnormal user or invasion when information immediately issued a message to the manager, this method in the current network security plays an important role. At present, anti-intrusion detection system is used to detect the main feature of this method is an expert in advance of the characteristics of the data set, so that the system is in a certain mode to detect intrusion data.

There are certain advantages in the application, the timely discovery of the invasion of information, but not be able to timely information to update the invasion, so for some emerging information will not recognize the invasion, during operation will often leak alarms and false alarms phenomenon. In addition, as the network's popularity in people's lives, network data are constantly increasing, so the audit record will be a lot of irrelevant information. This information will reduce the detection rate or lead to data overload. Web data mining system is mainly based on data association rules, categories and columns like sequence mode, intelligent analysis of the data by the law, so well established rules and exception monitoring model in intrusion detection system, through which kinds of programs to maximize the reduction in dealing with audit data on a priori knowledge discovery, which largely reduces the false alarm rate of the system.

The third one is Malicious code detection. In the anti-malware research, the most classic is the "signature" detection technology. But the signature detection technology has a fatal weakness, it can only detect known malicious code, malicious code for emerging it will not do anything. The use of Web data mining technology can effectively improve the quality and efficiency of malicious code detection. First, gather a large number of malicious code, malicious code library formed by adding some of the normal code in a large number of malicious code, and then divided into two parts, known as the training set, and the other part is called the test set; secondly, using a variety of algorithms, such as the rules of classification algorithms and Bayesian algorithms for training samples for training to correctly identify malicious code and normal code; Finally, the test set to evaluate their training. In general, the treated detect this malicious code, can achieve more satisfactory results.

The forth one is malicious spam detection. With the development of information technology, many newborn malicious code based on the emergence of e-mail attachments, and for this malicious code, the current common practice is to detect binding virus scanner via e-mail filter, and the virus scanner are based on the signature (signatures) to detect malicious code, for unknown malicious programs without a corresponding signature, thus the high cost of prevention, but less efficient. Built on Web data mining technology based on e-mail filtering system to e-mail for the detection of an object by analyzing a sample of the e-mail, get the most can distinguish whether it is a characteristic pattern of malicious e-mail, you can automatically discover newborn malicious programs, and then in this mode, based on the use of naive Bayes classifier and enhanced methods of machine learning and, ultimately, a mail filter to filter out malicious e-mail, get useful messages.

The fifth one is virus warning. With the popularity of the Internet, especially the prevalence of broadband networks, computer virus also to the network-oriented development, this virus is called worm. Traditional anti-virus techniques are based on the known virus signatures to identify and killing the virus, but the virus emerging on the powerless, so traditional antivirus technology with a lag. Virus warning provides feasibility, the use of Web data mining techniques to build a worm early

warning system that can connect behavioral anomaly detection network in real time, and thus able to find traces of worms, especially for emerging worm virus Web data mining technology has a role in early warning, allowing network administrators before this worm outbreaks can take appropriate measures to avoid big losses.

The sixth one is security assessment. Information security risk assessment is the basis for the work of information security protection management information system is an important part of risk management. Web data mining in the establishment of network information security risk assessment techniques based on the use of meta-search engine to retrieve information on the structure of the resulting process, allowing users to a large number of semi-structured risk assessment information to select Web, mining, integration , documentation, dig out the network information security risk assessment information and risk assessment in accordance with the characteristics of different information on the retrieved information to classify risk, in order to establish information security risk assessment information database, while constantly on the database information is updated, expanded, and then on this basis, establish a risk decision support system, providing strong support for the information network information security risk management.

Conclusion and Consideration

From the mental-anylasis of network information security and Web data mining techniques,we can know that web data mining technology is a key technology to enhance the performance of network information security. It can effectively improve network and information security.

Meanwhile, we also should know that technology applications in the prevention process to establish a network of information should be noted: Play the main role of workers, so mining results more accurate, more valuable; Protect the user's privacy. We should know that privacy mainly consider the following two points: First, no invasion of privacy for sensitive data such as ID number, name, address, etc. must be corrected and finishing in the original database; and second, the use of data mining algorithms for mining from the database information out of the invasion of privacy must also be excluded; Recognize the negative impact. Web data mining can be used by attackers to attack the network system for reasoning and aggregation attacks, stealing data, security breaches of the system.

Reference

[1] Iresearch; China highlights four individual network security market development trends.

[2] SEIFERT JW.Data mining and the search for security challenges for connecting the dots and data bases.Government Information Quarterly, 2004 (21):461-480.

[3] C.S.Tu,M.Y.Lu,Y.C.Lu.Web Summary of Data Mining [J].Computer Engineering and Application 2003,39 (10): 90-93.

[4] D.J.Li.Web data mining tool and its application in E-commerce[J].Microcomputer Applications, 2002 (7):180.