Research on the Application of Data Mining Technology in Computer Network Virus Defense

Huichuan Liu\textsuperscript{1,}\textsuperscript{*}, Dan Wang\textsuperscript{1}
\textsuperscript{1}Shandong Labor Vocational and Technical College, China

*Corresponding author e-mail: wangkan@sdltc.cn

Abstract. At present, the network virus defense is facing a more severe form, especially the emergence of various network viruses, which leads to the emergence of network security incidents in an endless stream, causing more serious damage and loss to related industries and fields. The traditional network virus defense measures have some defects, such as information lag, data update slow and backward technology, which have been difficult to deal with the real virus threat. Based on this, this paper first analyses the utilization status and advantages of data mining tech, then studies the utilization of network virus defense based on data mining, and finally gives the specific development strategy of data mining tech of computer network virus defense.

Keywords: Data Mining, Computer Network, Virus Defense

1. Introduction

With the iterative expansion of network info tech represented by computer, it has been widely and deeply popularized and studied in many fields, especially in the field of network virus defense, which greatly promotes the improvement of network virus defense level [1]. On the other hand, with the iterative maturity and progress of data mining tech, its utilization in network virus defense has made significant progress. At present, the network virus defense is facing a more severe form, especially the emergence of various network viruses, which leads to the emergence of network security incidents in an endless stream, causing more serious damage and loss to related industries and fields.

As a relatively new intelligent tech, data mining can effectively find out and defend the threats and attacks of network viruses represented by worms, so as to guarantee the security of network data and info [2]. With the help of data mining, we can analyze and judge the content of network data, and study the potential association between network data, so as to establish data premise for effective virus identification and mining. With the rapid popularization of computer network in various industries, it brings great convenience to the related fields [3]. At the same time, it also makes the related industries and fields face severe network virus threat and damage. In this context, the defense data of network virus need to be deeply mined to identify the weak points and hidden threats in the data.

In addition, computer network virus will cause great losses to computer users, and the threat and damage of network virus has a positive correlation with the degree of informatization. The traditional network virus defense measures represented by firewall have some defects, such as information lag,
slow data update and backward tech [4]. It is difficult to deal with the real virus threat. Only by making full use of data mining tech, can the security of network info transmission be better guaranteed.

In a word, under the realistic threat of computer network system security and Internet virus, the utilization of data mining tech can effectively deal with the realistic threat of network virus. And it can effectively identify the potential threat of data by mining the association relationship between data, and establish the core mechanism of data security protection, so as to achieve the ideal effect of virus protection. Therefore, it is of great practical value to study the utilization of data mining tech in computer network virus defense.

2. Utilization status and advantages of data mining tech

2.1. The connotation of data mining tech
Data mining refers to the tech of finding its rules from a large number of data, which is the integration of statistics, database tech and artificial intelligence tech. The tech automatically extracts several aspects of the content as shown in Figure 1 from the data, so as to realize the continuous improvement of the network virus prediction model. Secondly, data mining is a semi-automatic process; its main methods include data mining, association analysis, sequence pattern, classification, aggregation and anomaly detection. With the help of the analysis of historical data, this method can effectively predict the threat of network data and effectively mine its rules.

![Figure 1](image)

**Figure 1.** Data content extracted by data mining tech.

2.2. Anomaly detection of network data
As an important aspect of data mining, anomaly detection of network data is used to find patterns that are smaller than clustering, that is, objects that are significantly different from other data in the data set. Anomaly detection of network data can be effectively applied to the defense of network virus. The anomaly of network data is the unique data in the data set, which does not belong to clustering or background noise. The core elements of data mining algorithm include schema description language, schema evaluation and schema exploration. Among them, the schema description language hardens the knowledge types of data mining, its evaluation model reflects the knowledge model, and the schema exploration mainly focuses on the study of the schema parameter space and schema space. The algorithm and model characteristics of data mining system are shown in Table 1.

**Table 1.** The algorithm and model characteristics of data mining system.

| Integration               | Data model             | Features                                           |
|---------------------------|------------------------|----------------------------------------------------|
| Independent system        | Vector data            | Data mining as an independent utilization          |
| Data management system    | Continuous media data  | Integration with database and data warehouse       |
| Prediction model          | Semi structured data   | Oracle model system integration                    |
| Mobile system             | Computational model    | Combine with mobile data / various computing data  |
3. Utilization of network virus defense based on data mining

3.1. Classification and prediction of network virus

According to the training data set and class label attributes, the data mining algorithm constructs a data model to classify the existing data and to classify the new data. First, a model is built to describe the pre-determined data class set and concept set. The learning model can be provided in the form of classification rules, decision tree or mathematical formula, as shown in Figure 2 below. Secondly, the model is used to classify the unknown network security threats and evaluate the prediction accuracy of the model. For each test sample, the learning model class prediction of the sample is compared to prevent over adaptation to the data. In addition, through data preprocessing, the accuracy of network security threat object classification and prediction process is improved. For example, through data cleaning, correlation analysis and data transformation, noise can be eliminated or reduced, confusion in learning can be reduced, and data can be generalized to a higher level concept to realize the standardization of network security data.

![Figure 2. Scheduled data class set model.](image)

3.2. Cluster analysis of network virus

The utilization of data mining in clustering analysis of network virus can effectively understand the distribution of network data, and the data preprocessing steps are as follows: firstly, through clustering to find feature space to build index, in spatial data mining, detect and interpret clusters in space. Secondly, the quality of the clustering results lies in the similarity evaluation means and the specific execution of the clustering method. The evaluation criteria of the performance of the clustering method mainly lie in scalability, processing different types of attributes, discovering clusters of arbitrary shape, and the ability to handle data anomalies, so as to produce the clustering results that can meet the specified constraints.

In addition, at the level of data standardization, the average of absolute deviation and the standard measure are calculated. The average of absolute deviation is more robust than the standard deviation. Distance is used to measure the dissimilarity between two objects, and a weight is given according to the importance of each variable. Continuous ordinal variables have no units, the relative order of values is necessary, but the actual size is not important.

3.3. Association analysis of computer network virus data

Association rules mining of network security data is to find causal structures between item sets or object sets in network security data, relational data or other info carriers [5]. By describing the correlation rules between an item set and other item sets, all the rules are found with minimum support and credibility. The important step of association rule mining is to mine frequent sets, find out the item sets that fulfill the minimum support, and use the frequent sets to generate association rules.

In addition, at the level of multi-level association rules of network security data, the support of underlying items is usually low, and the rules of some specific layers may be more meaningful [6]. Transaction databases can be coded according to dimensions or layers, and can be shared for multi-
dimensional mining. At the level of mining multi-level association rules, mining multi-level association rules, that is, the association rules between different levels of cross or different hierarchical methods, is carried out by using the principle of uniform support or decreasing support between each level.

4. Specific development strategy of data mining tech for network virus defense

4.1. The feasibility of data mining tech utilization
Firstly, the utilization of data mining tech in computer network virus defense can effectively identify the security threats in network data, so as to strengthen security protection. In particular, the utilization of this tech is helpful to carry out the defense and protection of network virus from the overall level, and achieve the isolation of virus spread through the effective mining of virus propagation process and law. Secondly, the virus defense system based on data mining tech can effectively detect the scanning and intrusion of network threats represented by worms. In addition, by scanning the data source and building the database, it could establish the data preprocessing mechanism, effectively record the network spread virus, effectively record the potential network security events, and stimulate the start of network protection and defense mechanism.

4.2. Virus defense system module of data mining tech
The virus defense system module of data mining tech is shown in Figure 3 below. The decision module is mainly used to judge and compare the matching between the rule data and the result database. Secondly, the data collection module constructs the function information and data structure by capturing the data in the network. In addition, the data mining module is mainly used to generate analysis results with clear features and obvious rules. In addition, the rule base module is mainly in charge of the characteristic analysis of computer network virus. Finally, the data preprocessing module can effectively shorten the processing time between data mining and data analysis, and further enhance the ability of network virus identification and judgment.

![Figure 3. Virus defense system module of data mining tech.](image)

4.3. Internet virus defense process based on data mining
First of all, according to the data mining management rules to find the association among the system data, and clear the association rules in the network virus data. Secondly, to mine out the data with special value content, lay the premise of unknown virus search. In addition, through cluster analysis of the data, the relationship and typical distribution characteristics of the data are identified. Finally, by storing different data individuals in the predetermined classification, it could create a scientific data classification model, and rely on the content reflected by the data info to carry out further analysis and virus defense.

5. Conclusion
In summary, data mining can effectively identify and defend the threat and attack of network virus, so as to ensure the security of network data and information, achieve the analysis and judgment of
network data content, and mine the potential association between network data, so as to establish data premise for effective virus identification and mining. In this paper, through the research on the utilization status and advantages of data mining tech, the connotation of data mining tech and the advantages of anomaly detection are analyzed. Through the analysis of network virus defense utilization based on data mining, the classification and prediction, cluster analysis and association analysis of network virus are studied. Finally, the feasibility of the utilization of data mining tech and the utilization strategy of the virus defense system module of data mining tech are analyzed.

References
[1] Li Jiajia. On the utilization of data mining in computer network virus defense [J]. Network security tech and utilization, 2017, (08): 84 + 89.
[2] Liu Chunjuan. Utilization analysis of data mining tech in computer network virus defense [J]. Electronic test, 2014 (05): 83-85.
[3] Liu Qingyi. Utilization of data mining tech in computer network security virus defense [J]. Microcomputer utilization, 2018, 34 (04): 73-75.
[4] Sun He. Analysis of the utilization of data mining tech in computer network virus defense [J]. Jilin labor protection, 2016, (11): 28-29.
[5] Zhang Yan. Utilization of data mining tech in computer network virus defense [J]. Journal of Taiyuan City Polytechnic, 2016 (04): 174-176.
[6] Zheng Gang. Discussion on the utilization of data mining tech in computer network virus defense [J]. Information and computer (theoretical Edition), 2016, (03): 25 + 37.