APPLICATION OF ARTIFICIAL INTELLIGENCE IN COMPUTER NETWORK TECHNOLOGY UNDER THE BACKGROUND OF BIG DATA

L. Ma
Chongqing Youth Vocational & Technical College, Chongqing, 400712, China
Email: malcquy@126.com

Abstract: With the rapid development of society and technology, computer technology is also developing. Under the background of big data, the emergence of artificial intelligence is greatly beneficial to the life and work of people and moreover drives computer technology to be informationalized and intelligentized. The emergence and application of artificial intelligence enhances working efficiency and promotes the further development of many new technology fields in addition to enriching the life of people. Application of artificial intelligence in computer network can be classified to network security management, system evaluation and network management and artificial intelligence agent. This study firstly analyzed relevant theories of big data era and artificial intelligence, then designed the model structure of intelligent anti-spam in network security management, and tested the model. The results suggested that the intelligent anti-spam model had favourable filtering performance.

Keywords: Big data, Artificial intelligence, Computer network, Information filtering.

1. Introduction

The high-speed development of science and technology currently provides computer technology with a wide progress space. Moreover requirements on computer technology become higher and higher with the development of society [1, 2]. Under the background of big data, computer not only has the ability of data calculation and processing, but also can replace manpower to fulfil works based on its high-end orientation and intellectualization [3] to save time and manpower [4]. Application of artificial intelligence facilitates the life of people. Ahmad et al. [5] investigated the application of artificial intelligence in improving the service quality of computer network and proposed an artificial neural network model to dynamically control the pre-emption rate of on-going call in the network which starts using Quality of Service (QoS). The model maps network traffic parameters and desired operating pre-emption rate by network operator providing the best for the network under consideration into appropriate tuning parameter.

The simulation results suggested that the pre-emption rate closely matched the target rate.

Artificial intelligence contributes greatly to computer network security management. Hu et al. [6] put forward an intelligent hybrid spam filtering framework which could identify spam e-mail by analyzing the title of e-mail. Because of the high efficiency and extendibility, the framework was especially applicable to giant e-mail server. They extracted five features from the title of e-mail, i.e. the field of addresser, the field of target, the field of X-mail, the IP address of addresser server and e-mail subject. E-mail subject was digitalized using n-gram based algorithm to obtain better performance. The experimental results demonstrated that the framework based on random forest algorithm had favourable accuracy, recall rate, preciseness and F-measure. After the addition of MetaCost framework, the model operated stably and generate small cost in cost-sensitive conditions.

This study aims at analyzing the application of artificial intelligence in computer network through designing and testing an intelligent anti-spam e-mail model, which lays a theoretical basis for the development of artificial intelligence.

2. Overview of Big Data Era and Artificial Intelligence

2.1 Characteristics of Big Data Era

Big data refers to a large number and various types of data sets. The characteristic of big data era can be summarized as 4Vs, i.e. volume, velocity, variety and veracity. Volume means large data volume. Capacity of data which accumulate through network behaviours is large, more than 10 TB generally. Velocity means rapid data transmission and processing speed. Data on the network increases

explosively; hence the data updating and processing capability must be improved to satisfy the requirement of production and life. Variety means varied types of data. Big data contains multiple types of data, structured data and unstructured data in simple classification [7]. Veracity means high veracity of data.

Traditional data are replaced by new data, which improves veracity and security and ensures data not to be influenced during transmission and processing.

2.2 Application of artificial intelligence

Artificial intelligence is a comprehensive technology which integrates multiple subjects.

Computer replaces human to do complex and danger works using the ability of simulating the thinking of human based on the codes input during computer programming. The effective application of artificial intelligence can fundamentally reduce working time and improve working efficiency. Under the background of big data era, the application of artificial intelligence in computer network mainly includes the following three aspects.

Network security management: First is intrusion detection. As the core component of firewall, intrusion detection can ensure the security, reliability and integrity of computer network resources [8]. Intrusion detection can monitor the operation state of network and protect network on the premise of network performance [9].

The second aspect is intelligent firewall. Intelligent firewall system can collect, analyze and process data using intelligent identification technology and intercept, which improves the security grade of information [10], prevents the transmission and diffusion of virus, effectively prevents and solves network security problems, avoid external attacks, and enhances the effectiveness of network security management.

The third aspect is intelligent anti-spam e-mail. Intelligent anti-spam e-mail system monitors, scans and identifies user mailbox and e-mail to extract spam e-mail, ensure information security, and classify spam e-mail [11].

System evaluation and network management: Accelerating the intelligent management progress can effectively enhance network management efficiency and quality. The comprehensive management of network can be realized by establishing comprehensive management system based on expert knowledge base and solution techniques [12]. Expert system which integrates the knowledge of experts in some field can analyze and solve professional problems in relevant field. The emergence and application of expert system can further improve the level of network management.

Artificial intelligence agent management: Artificial intelligence agent can analyze and process information data based on the knowledge of different agents, search information based on user defined, and transmits data to the specified position, which can achieve more humanized and intelligent services. Moreover, artificial intelligence agent technology has autonomy and learns ability; it can more effectively fulfil the tasks assigned by users through learning evolution [13].

3. Design of Intelligent Anti-Spam E-Mail System

3.1 Flow of anti-spam e-mail system

Anti-spam e-mail system in this study was based on Linux platform firewall. The intelligent firewall system has functions of filtering spam e-mails, websites and text messages. Only filtering of spam e-mail was discussed in this study.

The system intercepted data package of an e-mail from network, analyzed the data package of the e-mail, and identified the content of the e-mail. When there was decadent content in the content of e-mail, the e-mail should be abandoned. The detailed work flow is shown in Figure 1.

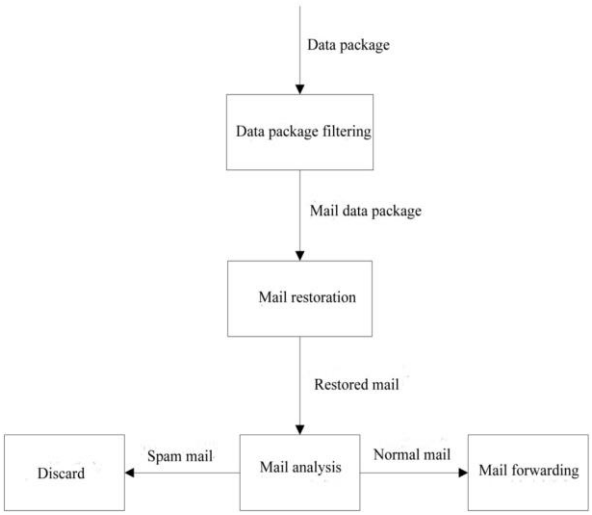

Figure 1: The work flow of anti-spam e-mail system

3.2 E-mail protocol analysis

SMTP protocol: Simple Message Transfer Protocol (SMTP) [14] has two working modes, sending SMTP and receiving SMTP. The specific working process was as follows. Client connected 25 port of SMTP server through transmission control protocol (TCP). To identify the identity of addresser, client sent HELO to server and then MAIL. Server sent OK as a response after receiving the command and then prepared to receive. Client sent RCPT. Then server should display whether it was willing to receive e-mail. E-mail was sent out using DATA after negotiation. Finally QUIT was used as an end.

The specific implications of some commands are shown in Table 1.

Table 1. Table of relevant SMTP commands

SMTP command	Implications of commands
HELO	Client sends the command to establish connection with SMTP server and sends E-mail address to SMTP server
MAIL	Client sends the name of addresser to SMTP server
RCPT	Client sends the name of receiver to SMTP server
DATA	Client sends mail content to SMTP server
QUIT	Terminating the connection between client and SMTP server after receiving the response of OK replied by SMTP.

POP3 protocol: Post office protocol-version 3 (POP3) is mainly used for supporting remote management of e-mail on server by client. The specific working principle was as follows. Client connected with 110 port via TCP and sent mail user name and password to POP3 for authentification of user using USER and PASS. Then it requested server to send the statistical data using STAT, listed the number of e-mails on server using LIST, and received e-mail using RETR. After that, e-mail in mail server was labeled as deleted state using DELE. Then QUIT was sent to delete e-mails which were labeled as deleted.

Table 2 Table of relevant POP3 commands

POP3 command	Implications of commands
USER	Processing user name
PASS	Processing user password
STAT	Request server to send statistical data about mailbox such as number of mails and total number of bytes
LIST	Sending back the number of mails and the size of each mail.
RETR	Sending back all the texts of e-mails with parameter identifier
DELE	Labeling mails with parameter identifier as deleted
QUIT	Delete mails with labels in server and quit.

3.3 Design of intercepting, redirection and restoration of e-mail

Intercepting: The content of e-mail should be acquired firstly before the intercepting of spam e-mails. As e-mails based on SMTP and POP3 were established on TCP connection and communicated using relevant protocol port, 25 port of SMTP or 110 port of POP3 was monitored to obtain data package and relevant commands were analyzed to obtain important information such as sender, receiver, text content and attachment. The obtained domains were filtered according to different processing rules.

Redirection: The system mounted Hook function which could process mail data package in Forward node and set the processing action as NF-QUEUE. Then IP data package whose source and destination port were corresponding protocol port number were inserted in Netfilter_queue to reorient from kernel mode to user mode program. Moreover data packages except mail data package were transmitted unconditionally. Matching stopped after database matched successively after satisfying matching conditions at the first time. After data processing, legitimate data packages flew to network, while illegitimate data packages were abandoned.

Restoration: IP data package fragment should be recombined before restoring mails based on data package. Firstly whether IP data package was fragmented was determined according to the fragment mark field extracted by header and deviant. If it was confirmed as fragmented, then IP data package was stored using sk_buff structure. The storage position was designed as the position after sk_buff structural data. Data pointer pointed at the first letter of data. The position of the current fragment in all data package fragments was positioned. Whether the fragments of the same data package arrived was checked. If it was not, the fragments were controlled; if it was, then all the fragments were integrated to restore mail. Finally the interactive content and mail were restored via protocol and according to the data flow of mail sending and receiving.

3.4 Bayesian classification algorithm

Figure 2 suggested that the determination on spam e-mail was based on Bayesian classification algorithm. Statistical algorithm based on Bayesian rules is one of the most effective anti-spam e-mail technologies currently. Moreover it has autonomic learning ability because of the addition of neural network method. Whether there were spam words and whether an e-mail was spam were determined by comparing the occurrence frequency of the same words in the previously received spam e-mails and legitimate e-mails [15].

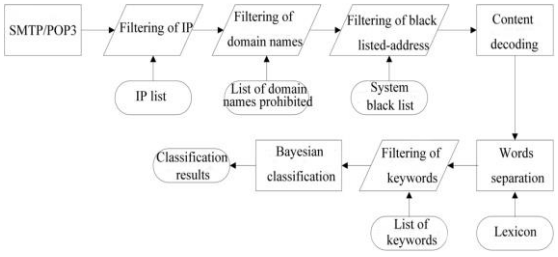

Figure 2: The detailed filtering processes of SMTP and POP3

$$a(w) = \frac{\text{Number of spam e - mails containing word } w}{\text{Number of all spam e – mails}} \quad (1)$$

$$b(w) = \frac{\text{Number of non – spam e – mails containing word } w}{\text{Number of all spam e – mails}} \quad (2)$$

Then the probability of whether an e-mail containing some word was spam was:

$$c(w) = \frac{a(w)}{a(w) + b(w)} \quad (3)$$

The improved algorithm was

$$d(w) = \frac{(x * y) + (z * c(w))}{x + z} \quad (4)$$

where x stands for the concentration of background information, y stands for the hypothetical probability of background information, and z stands for the total number of e-mails containing w.

Probability of whether an e-mail was spam could be calculated based on the above probability. The formula was:

$$S = C^{-1}(-2\ln \prod_{\omega} d(w), 2z) \quad (5)$$

where C-1() stands for inverse chi-square function.

f(w) could be replaced by (1-f(w)) to calculate the probability whether an e-mail was non-spam.

Then the probability of whether an e-mail was spam was:

$$Q = \frac{1 + S - P}{2} \quad (6)$$

Finally the calculation results were determined. If the result was close to 0, then it was determined as spam e-mail. If the result was close to 1, then it was determined as non-spam e-mail. If the result was close to 0.5, then it was uncertain.

3.5 System testing and results

Testing environment: mail sender, mail receiver and mail gateway were connected together by an interaction machine; moreover mail gateway was connected with mail server.

Testing method: for the first time, mail sender sent 500, 1000, 1500 and 2000 spam mails to mail receiver respectively; for the second time, mail sender sent 500, 1000, 1500 and 2000 non-spam mails to mail receiver.

Testing results: The identification results of spam e-mails and non-spam e-mails were shown in Table 3.

Table 3 demonstrates that the identification rate of non-spam e-mails was higher than that of spam e-mails; the identification rate of non-spam e-mails increased steadily, while the identification rate of spam e-mails fluctuated with the changes of number of e-mails. The identification time of the two kinds of e-mails became increasingly shorter with the increase of number of e-mails, suggesting the identification speed became higher with the increase of number of e-mails; the identification speed of spam e-mails was high when the number of e-mails was small, and the identification speed of non-spam e-mails was high when the number of e-mails was large.

Table 3 The identification results

Identification results		Number of e-mails			
		500	1000	1500	2000
Identification degree (%)	Spam e-mail	92.4	91.7	93.6	92.9
	Non-spam e-mail	97.3	97.7	98.3	98.7
Identification time (ms)	Spam e-mail	540.3	531.9	529.4	527.9
	Non-spam e-mail	541.1	532.5	528.5	520.3

4. Conclusion

Artificial intelligence is the product of network, communication and computer technology. Under the background of big data era in which data increase explosively, methods which can process data effectively and efficiently are needed. The emergence of artificial intelligence effectively relieves the problem of insufficient ability of information processing in the current stage. Moreover artificial intelligence can rapidly process information on the premise of ensuring information security. In this study, relevant theories of artificial intelligence was analyzed, the anti-spam e-mail system was designed. Mail data were obtained through intercepting, redirecting and restoring SMTP and POP3 port data packages, and then the mails were filtered using Bayesian algorithm. The testing results suggested that the system had high identification rate and could filter and intercept spam e-mails. The application of artificial intelligence in computer network technology can not only solve the problems such as insecurity and slow data response in the current network technology, but also is beneficial to promote social development.

References

[1] V.I. Talanin: “Informatization of modern society: the history, ideology, technologies”, Science & Technology, Vol. 3 No. 2A, pp. 48-59, 2013.

[2] O. Seregina: “The development of science and technology as a factor contributing to changes in modern legal systems”, Law & Modern States, Vol. 3, pp. 9-16, 2013,

[3] M. Mansourvar, M.A. Ismail, S.A. Kareem, R.G. Raj, F.H. Nasaruddin, R. Mahmud, R. Abdullah and N. Idris: “A Computer-based system to support intelligent forensic study”, Fourth International Conference on Computational Intelligence, Modelling and Simulation, pp. 117-119, 2012.

[4] N. Li, N. Matsuda, W.W. Cohen, and K.R. Koedinger: "Integrating representation learning and skill learning in a human-like intelligent agent", Artificial Intelligence, Vol. 219, pp. 67-91, 2015.
[5] I. Ahmad, J. Kamruzzaman and D. Habibi: "Application of artificial intelligence to improve quality of service in computer networks", Neural Computing & Applications, Vol. 21 No. 1, pp. 81-90. 2012,
[6] Y. Hu, C. Guo, E. Ngai, S. Chen and S. Chen: "A scalable intelligent non-content-based spam-filtering framework", Expert Systems with Applications, Vol. 37 No. 12, pp. 8557-8565, 2010.
[7] C. Ollivier-Gooch, L. Diachin, M.S. Shephard, T. Tautges, J. Kraftcheck, V.J. Leung, X.J. Luo and M. Miller: "An interoperable, data-structure-neutral component for mesh query and manipulation", ACM Transactions on Mathematical Software, Vol. 37 No. 3, pp. 1-28, 2010.
[8] C. Modi, D. Patel, B. Borisaniya, et al. "Review: A survey of intrusion detection techniques in Cloud", Journal of Network & Computer Applications, Vol. 36 No. 1, pp. 42-57, 2013.
[9] C. Panos, C. Xenakis, P. Kotzias, et al. A specification-based intrusion detection engine for infrastructure-less networks", Computer Communications, Vol. 54 No. C, pp. 67-83, 2014.
[10] H. Jo, S. Kim and D. Won: "Advanced information security management evaluation system", Ksii Transactions on Internet & Information Systems, Vol. 5 No. 6, pp. 1192-1213, 2011.
[11] G. Sakkis, I. Androutsopoulos, G. Paliouras, V. Karkaletsis, C.D. Spyropoulos and P. Stamatopoulos: "Stacking classifiers for anti-spam filtering of e-mail", International Journal of Production Research, Vol. 52 No. 19, pp. 5857-5879, 2012.
[12] A. Martín, C. León and F. Biscarri: "Intelligent Integrated Management for Telecommunication Networks", International Journal of Advancements in Computing Technology, Vol. 2, pp. 158-171, 2010.
[13] Y. Sandamirskaya and M. Burtsev: "NARLE: Neurocognitive architecture for the autonomous task recognition, learning, and execution", Biologically Inspired Cognitive Architectures, Vol. 13, pp. 91-104, 2015.
[14] S. Al-Fedaghi and A. Mohamoud: "Conceptual description of simple mail transfer protocol", Far East Journal of Electronics & Communications, Vol. 11 No. 2, pp. 113-133, 2013.
[15] T. Subramaniam, H.A. Jalab and A.Y. Taqa: "Overview of textual anti-spam filtering techniques", International Journal of the Physicalences, Vol. 5 No. 12, pp. 1869-1882, 2010.