Research on trade data encryption of tobacco enterprises based on adversarial neural network

Zhang Yi

Accepted: 21 October 2021 / Published online: 1 December 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
In order to overcome the problems of long encryption process, low data security and poor anti-attack rate of traditional tobacco enterprise trade data encryption methods, a trade data encryption of tobacco enterprises based on adversarial neural network is proposed. This method optimizes the traditional neural network by generating countermeasure network, so as to form adversarial neural network. In the adversarial neural network, the encryption processing of tobacco enterprise trade data is completed through data feature classification, design of tobacco enterprise trade data encryption protocol and data encryption channel. The experimental results show that the encryption process of this method takes between 6 s and 20 s, and the encryption efficiency is high, and this method can effectively scramble the original arrangement of the data, so as to effectively hide the effective information in the tobacco enterprise trade data, improve the anti-attack rate of the data, and effectively improve the security of the data.

Keywords Tobacco enterprise trade data · Adversarial neural network · Data classification · Encryption protocol · Data encryption

1 Introduction
With the steady progress of the reform and development strategy of the national tobacco industry, tobacco enterprises have gradually become bigger and stronger through combination, reorganization and merger. The industrial organization system structure shows an obvious trend of integration and centralization. This change in the mode of production, operation, organization and management of the industry puts forward higher-level requirements for the construction of tobacco informatization. This requirement is mainly reflected in the use of informatization to support industry centralized management and auxiliary decision support. China’s tobacco business has a significant production and market scale after years of information building and fast market expansion, and has made good contributions to national, local finance at all levels, and national economic construction. A series of tobacco trade centers have established relatively mature business systems, including three-level batch business system, website system, sales system, etc., these business systems have produced a large amount of data and are gradually developing toward mass. These data are very detailed and stored in different computer systems. The transmission and storage techniques of this data have changed dramatically as electronic information and the Internet have evolved (Kumar & Bhatt, 2020; Sharma & Jinwala, 2021).

However, because the trade data of tobacco enterprises are easy to copy and spread, it will bring many security threats to commerce (Pérez-Resa et al. 2020). At the same time, in order to meet the growing industry demand and the continuous progress of network and information technology, the secure transmission of tobacco enterprise trade data in network and insecure channel has become an urgent problem to be solved (Ikeda et al. 2019). Data security research involves many contents, including security architecture, cryptographic protocol, security theory, security performance analysis, etc., among them, encryption technology is one of the key technologies to ensure information security.
There have been many significant advances in the development of artificial intelligence in recent years. In some cases, AI can learn on its own and perform some tasks better than humans. The key data feedback encryption technique based on scheduling model is developed, for example, in (Yi, 2021). The key encrypted data are segmented using the 3GR encrypted string segmentation technique in this approach, and then an AEHA complete authentication model is built. The truncated string is optimized to complete data encryption on the basis of accelerated authentication. In (He, Zhao, & Cheng, 2021), a data homomorphic encryption method based on improved RSA algorithm is designed. This method first uses DES algorithm to encrypt plaintext and RSA algorithm to encrypt key, then calculates ciphertext sequence in plaintext and ciphertext space through addition homomorphism process, and then obtains plaintext data through corresponding decryption operation (Muhammad Talha et al. 2020a, b).

However, it is found in the practical application that the traditional data encryption method has the problems of the encryption process time, data security and anti-attack rate. Therefore, this study designed a tobacco enterprise trade data encryption method based on adversarial neural network (Muhammad Talha, 2021).

2 Adversarial neural network analysis

2.1 Traditional neural network

Neural network is a computer model structure with a wide range of applications, which contains a large number of upper and lower related units, and these basic units are named neurons (Elhadad, 2020). These neurons are divided into at least two layers according to the logical structure of the network, and different layers are interconnected through weighted connections. These weighted parameters largely determine the characteristics of the two associated neurons (Yoshida & Fujino, 2020). In the structure of neural network, basically each neuron is connected with the neuron at the next level. After the data are input from the input layer, the data processing results are output in the output layer through the action of multiple neurons. The intermediate layer sits between the input and output layers, and its job is to build a complicated network function that completes data processing.

Not only can a neural network do self-repair, but it also has a high fault tolerance, preventing data vulnerabilities and aberrant nodes. In addition, the neural network model can also infer the correlation between different data nodes, which enables it to have the ability of non-linear data processing (Sartakhti & Jalili, 2019).

However, although the neural network has so many advantages above, it relies too much on the performance of the system hardware itself, that is, it cannot explain the network behavior, nor can it choose the optimal network structure, nor can it set the training time in the process of network training (Saračević et al. 2020). These disadvantages also restrict the development of neural network to a great extent (Muhammad Talha et al. 2021).

2.2 Adversarial neural network

Based on the aforementioned neural network research, this study improved the conventional neural network by creating adversarial networks, resulting in adversarial neural networks that may be used to encrypt trade data for cigarette companies.

Generative adversarial network can effectively distinguish the distribution of high-dimensional and complex data, and it applies adversarial method to mine the generative algorithm of data distribution (Vuppala et al. 2020). In generative adversarial network, there are two main structures: generative module S and discriminant module P. The purpose of generating modules is to generate false data according to the implicit variable y, while the purpose of discriminating modules is to form real data space based on the input judgment parameters. The discriminant module can output a specific value if it is able to determine that the input information is from a real data space. Because the generative module and the discriminant module compete with each other, they are antagonistic. The network parameters can be optimized in this adversarial learning way, and the process is as follows:

\[
\min_S \max_P V(S, P) = \min_S \max_P p_{x-data} \log D(x) + p_y \log(1 - D(y))
\]

(1)

In formula (1), \(p_{x-data}\) and \(p_y\), respectively, represent the distribution probability of real data and the distribution probability of the hidden variable \(y\) in the data space. \(V(S, P)\) is a binary cross entropy function. The generating module maps the data into the real data space, and then the discriminant module determines whether the output data \(x\) is a fake sample or a real sample.

The real sample \(x\) identified through the above process is transmitted to the input layer as the input information of the neural network. From the point of view of data property classification, \(V(S, P)\) belongs to a kind of objective function. If \(x\) is real data, the discriminant module maximizes its output; If \(x\) is the pseudo data formed in the generated module, identify the module to minimize its output. Therefore, formula (1) will yield at least \(\log(1 - D(y))\) entries. In this process, the generating module tries to maximize the output information of the
3 Encrypt tobacco enterprise trade data based on adversarial neural network

By inputting the trade data of tobacco enterprises into the adjunctive neural network designed above, real data classification results will be obtained. On this basis, the obtained data will be encrypted to ensure the security of trade data of tobacco enterprises.

3.1 Characteristic classification of trade data of tobacco enterprises

The Bayes theorem, which is an essential theory for judging the connection between prior probability and posterior probability, is first used to categorize the features of tobacco business trade data in this article. When judging the probability of an event through known information, the obtained probability is prior probability, and the posterior probability can be obtained by deducing prior probability (Baagyere, Agbedemnab, Qin, Daabo, & Qin, 2020). Bayes’ theorem is as follows:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$  \hspace{1cm} (2)

In formula (2), $P(A)$ and $P(B)$, respectively, represent the prior probabilities of event $A$ and event $B$, and $P(A|B)$ and $P(B|A)$, respectively, represent the posterior probabilities of event $B$ and event $A$ when event $A$ and $B$ are known.

Naive Bayes classifier makes full use of Bayes’ theorem, which consists of three stages: preparation, training and calculation. In the preparation stage, the attributes of the data are divided; in the training stage, the samples are trained by the attributes of the divided samples; in the calculation stage, the probability of the occurrence of samples of different categories and the conditional probability of the occurrence of different data attributes in samples of different categories are calculated. On this basis, the final data feature classification result is obtained by comparing with the probability obtained in the training stage.

Suppose there is trade data set $C$ of tobacco enterprises, and $C = \{C_1, C_2, \cdots, C_n\}$, $n$ represents the data category in the set, and $A = \{a_1, a_2, \cdots, a_m\}$ represents the samples to be classified with the number of attributes $m$. The naive Bayes classification algorithm is adopted to obtain the maximum value of $P(C_n|A)$, and the process is as follows:

$$P(C_n|A) = \frac{P(A|C_n)P(C_n)}{P(A)}$$  \hspace{1cm} (3)

If $P(A)$ of each random classification is in the same state, only the largest molecule can be analyzed in the classification process. All attribute sets in sample set $A$ are in independent states, then formula (3) is transformed into the following form:

$$P(A|C_n) = \prod_{i=1}^{m} P(A_m|C_n)$$  \hspace{1cm} (4)

If sample set $A$ is at $i = q$, $P(A|C_n)$ has the maximum value, then the target classification is $q$.

3.2 Tobacco industry trade data encryption protocol

The results of feature classification are obtained by using known data information through naive Bayes, and then applied to the subsequent encryption process. In this paper, data encryption protocol and encryption channel are designed to complete encryption design.

To begin, the chaotic key protocol control process is comprised of full data encryption arithmetic coding and quantitative parameters in the optimization process, as well
as the sequence to encrypt the data set. \( X = \{x_1, x_2, \ldots, x_i, \ldots, x_n\} \), where \( x_i (1 \leq i \leq n) \) is the characteristics of the \( i \)-th a section coding sequence, the characteristics of the trade data distribution in tobacco enterprise is domain, the fuzzy characteristics of the data encryption process sequence are defined as follows:

\[
k_{x_i} = \frac{x_i(k + 1)}{p k}
\]

In formula (5), \( k_{x_i} \) represents the private key of the data sender (Masuda et al. 2020). On this basis, the spatial quantization coding method is adopted to reconstruct the data features, and the extended ciphertext sequence of the tobacco enterprise trade data to be encrypted is \( X(n) = x_1 + x_2 + \ldots + x_n \). Then, the proxy reschedule method is used for the block matching of sparse data features, and the block matching model is obtained as follows:

\[
f_{k_{x_i}} = \left[ s_{k_{x_i}} \right]_{x_i} C k_{x_i}
\]

In the block matching model, the convergence control method is adopted to implement linear random control on the trade data of tobacco enterprises, and the decryption key is obtained as follows:

\[
S = \sum_{i} f_{k_{x_i}} \sigma_{k_{x_i}}
\]

In formula (7), \( \sigma_{k_{x_i}} \) represents the similarity probability of ciphertext distribution in the trade data set of tobacco enterprises. Under the control of the private key, the extended key protocol to encrypt the trade data of tobacco enterprises is:

\[
Z(k) = \frac{f_{k_{x_i}}}{Sp}
\]

In formula (8), \( p \) represents the fuzzy constraint parameter in the data encryption process.

### 3.3 Data encryption channel design

Traditional data encryption methods only encrypt data within the local network, ignoring the encryption process in data interaction. Aiming at this problem, the design of data encryption channel is added in this paper. The data encryption environment is measured in this section based on the computing node big data environment’s efficient analysis and processing capabilities.

The conventional data channel state is determined by the runout frequency and runout mode of data node. The influence of multiple signal frequencies is involved (Vengala et al. 2020). The traditional channel frequency is static and its anti-attack is poor. This research dynamically extracts the features of various signals in the channel, canceling the appropriate wave frequency, based on the findings of big data analysis, in order to accomplish the goal of characterizing the channel environment. On this basis, the channel after quantization is encrypted and isolated, and the encryption layer and channel are fused together by joint optimization algorithm. The joint optimization process is as follows:

\[
G(u) = \int_{1}^{\hat{v}} u(h)^{\hat{v}} du
\]

In formula (9), \( u \) represents the channel frequency; \( h \) represents the synthesis degree of the encryption layer; \( \hat{v} \) represents the data node coefficient. In formula (9), the value of synthesis degree of encryption layer is related to the encryption quality of the whole data channel.

The larger the value of synthesis degree of encryption layer is, the more secure the data channel is. The value range of synthesis degree of encryption layer is dynamically allocated by the hidden layer of adversarial neural network according to data needs. The data encryption channel structure is shown in Fig. 2.

---

![Diagram](image-url)

**Fig. 2** Schematic diagram of data encryption channel structure

© Springer
At this point, the design of the encryption method of tobacco enterprise trade data based on adversarial neural network is completed.

4 Experiment and result analysis

4.1 Experimental design

In order to verify the practical application performance of the tobacco enterprise trade data encryption method designed above based on adversarial neural network, the following simulation experiment is designed using Matlab platform for verification.

The experiment uses a cigarette company’s trade database as the experimental object, and trade big data produced during the company’s operations as the experimental sample. The enterprise database has a total of 20 GB of data, which is divided into 15 data sets. The experimental environment was configured with two computers, both of which were configured with CPU I5 6200 (main frequency 3.2 Hz, memory 8G). The other experimental environments and parameters were shown in Table 1.

4.2 Preliminary results

To begin, the suggested technique is used to encrypt trade information data in the enterprise database, which is utilized to test the proposed method’s application impact. Before and after encryption, the data output histogram is shown in Fig. 3.

By analyzing the results shown in Fig. 3, it can be seen that the distribution frequency range of trade information data in the enterprise database is effectively encrypted after the application of the proposed method, indicating that the proposed method has good encryption performance and high data security.
Fig. 5 Comparison of ASCII value distribution of data before and after encryption by different methods
4.3 Compare the experimental results and analysis

Order to avoid the singularity, the experimental results in (Azmi & Kusumaningrum, 2019) design feedback scheduling model based on key data encryption method and the design in (He et al. 2021) data homomorphic encryption method based on RSA algorithm for comparison method, from the encryption process takes three angles, data security, and resistance to attack rate, launches the performance verification together with the method in this paper.

4.3.1 Time verification of different methods of encryption process

Through several iterations, the encryption process time of the method in this paper, method (Azmi & Kusumaningrum, 2019) and (He et al. 2021) method were counted, and the results obtained were shown in Fig. 4.

The analysis of the variation trend of line segments in Fig. 4 shows that with the constant change of the number of experimental iterations, the encryption process time of the three methods also changes accordingly. Among them, the encryption process time of the (Azmi & Kusumaningrum, 2019) method varies between 28 s and 54 s, the encryption process time of the (He et al. 2021) method varies between 34 s and 46 s, while the encryption process time of the method in this paper varies between 6 s and 20 s, which is significantly less than the two traditional methods. Therefore, it takes less time to verify the encryption process of the system in this paper, indicating that the method in this paper has higher encryption efficiency.

4.3.2 Different methods of data security verification

One of the most efficient indications of the application impact of data encryption is security. As a result, after using the technique in this article, the distribution of standard code values for information exchange (ASCII) was utilized as an indication in the experiment to evaluate data security., (Azmi & Kusumaningrum, 2019) method and (He et al. 2021) method. The experimental results are shown in Fig. 5.

Through the analysis of the results shown in Fig. 5, it can be seen that the proposed method can effectively scramble the original arrangement of data after encrypting the trade data of tobacco enterprises, thus effectively hiding the effective information involved in the original trade data of tobacco enterprises and improving the security of data. The two conventional techniques, however, failed to obfuscate the data’s original layout, and the attacker was still able to recover some of the delay information from the encrypted data. This demonstrates that the technique in this article has a substantial encrypting impact, indicating that its security is higher (Muhammad Talha, Azeem, Sohail, Javed, & Tariq, 2020).

4.3.3 The test of data anti-attack rate by different methods

In order to verify the data anti-attack rate of different methods, in this part of the experiment, three attack modes of data tampering attack, forgery attack and injection attack are simulated, and the anti-attack times and anti-attack rate of data subjected to different external attacks after encryption by this method, (Azmi & Kusumaningrum, 2019) method and (He et al. 2021) method are counted. The results are shown in Table 2.

Table 2 shows that the trade data of tobacco enterprises encrypted using this method can effectively withstand three
types of attacks, with the highest anti-attack rate of 99.50 percent, demonstrating that the trade data of tobacco enterprises encrypted using this method has good security performance. However, the anti-attack rate of the tobacco enterprise trade data encrypted by the (Azmi & Kusumaningrum, 2019) method and (He et al. 2021) method is significantly lower than that of the method in this paper. The above experimental results show that the proposed method can effectively ensure that the encrypted trade data of tobacco enterprises can resist the risk of information disclosure caused by external attacks, and the anti-attack ability of data is significantly improved.

5 Conclusion

Traditional data encryption techniques have disadvantages, such as inadequate encryption length, data security, and anti-attack rate. This study designed a new trade data encryption method for tobacco enterprises based on adjunctive neural network. According to the experimental results, the proposed method not only has high encryption efficiency, but also can effectively scramble the original data arrangement, so as to effectively hide the effective information involved in the original trade data of tobacco enterprises, and improve the anti-attack ability of the data, proving that it has achieved the design expectation.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Azmi SD, Kusumaningrum R (2019) Relevance feedback using genetic algorithm on information retrieval for indonesian language documents. J Inf Syst Eng Bus Intell 5(2):171–182
Baagtery EY, Agbedemmab PA-N, Qin Z, Daabo MI, Qin Z (2020) A multi-layered data encryption and decryption scheme based on genetic algorithm and residual numbers. IEEE Access 8:100438–100447
Chabanne H, Despiegel V, Guiga L (2020) A protection against the extraction of neural network models. arXiv preprint arXiv:12782
Elhadad A (2020) Data sharing using proxy re-encryption based on DNA computing. Soft Comput 24(3):2101–2108
He W, Zhao L, Cheng L (2021) A homomorphic encryption method for power data based on improved paillier algorithm. Paper presented at the International Conference on Artificial Intelligence and Security
Ikeda K, Sato Y, Koyama O, Yamada M (2019) Two-dimensional encryption system for secure free-space optical communication of time-series data streams. Electron Lett 55(13):752–754
Kumar P, Bhatt AK (2020) Enhancing multi-tenancy security in the cloud computing using hybrid ECC-based data encryption approach. IET Commun 14(18):3212–3222
Masuda H, Nakai T, Yoshida K, Kubota T, Shiozaki M, Fujino T (2020) Black-box adversarial attack against deep neural network classifier utilizing quantized probability output. J Signal Process 24(4):145–148
Pérez-Resa A, García-Bosque M, Sánchez-Azqueta C, Celma S (2020) A new method for format preserving encryption in high-data rate communications. IEEE Access 8:21003–21016
Saračević MH, Adamović SZ, Miškovic VA, Elhoseny M, Maček ND, Selim MM, Shankar K (2020) Data encryption for Internet of Things applications based on catalan objects and two combinatorial structures. IEEE Trans Reliaib
Sartakhli JS, Jalili S (2019) On the computational power of the light: A plan for breaking data encryption standard. Theoret Comput Sci 773:71–78
Sharma D, Jinwala D (2021) Simple index based symmetric searchable encryption with result verifiability. Front Comp Sci 15(2):1–3
Talha M, Sohail M, Tariq R, Ahmad MT (2021) Impact of oil prices, energy consumption and economic growth on the inflation rate in Malaysia. Cuadernos De Economı´a 44(124):26–32
Talha M, Azeeem S, Sohail M, Javed A, Tariq R (2020) Mediating effects of reflexivity of top management team between team processes and decision performance. Azerbaijan J Educ Stud, 690
Talha M, Tariq R, Sohail M, Tariq A, Zia A, Zia M (2020) Review of international geographical education ISO 9000: (1987–2016) a trend’s review Review of International Geographical Education Online, 10
Talha M (2021). Financial statement analysis of atlas honda motors, indus motors and pak suzuki motors (EVIDENCE FROM PAKISTAN). Ilkogretim Online, 20(4)
Vengala DVK, Kavitha D, Kumar AS (2020) Secure data transmission on a distributed cloud server with the help of HMCA and data encryption using optimized CP-ABE-ECC. Clust Comput 23(3):1683–1696
Vuppala A, Roshan RS, Nawaz S, Ravindra J (2020) An efficient optimization and secured triple data encryption standard using enhanced key scheduling algorithm. Procedia Computer Science 171:1054–1063
Yi, Z. (2021). Research on trade data encryption of tobacco enterprises based on adversarial neural network
Yoshida K, Fujino T (2020) Countermeasure against backdoor attack on neural networks utilizing knowledge distillation. J Signal Process 24(4):141–144

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.