WILEY | Hindawi

Research Article

Design of Shared Internet of Things System for English Translation Teaching Using Deep Learning Text Classification

Lin He,[1] **Jiaqi Guo,**[2] **and Jiaxin Lin** [3]

[1]*School of Foreign Languages, Southwest Medical University, Luzhou 646000, China*
[2]*Graduate School, The University of Melbourne, Victoria 3010, Australia*
[3]*School of Foreign Studies, Northwestern Polytechnical University, Xi'an 710129, China*

Correspondence should be addressed to Jiaxin Lin; jasonlin@gdufe.edu.cn

The purpose is to adapt to the current social development and promote the English translation teaching reform. Based on the theories of deep learning (DL), text classification (TC), and the Internet of Things (IoT), this work analyzes the current situation of English translation teaching. Additionally, 100 text categories are selected from the English text corpus of Northwestern Polytechnic University as the research objects. The data are classified by the DL-based TC method and analyzed by introducing the simulated annealing algorithm. Finally, the storage and security performance of the shared IoT system are described. The results show that the proposed TC method can overcome the performance loss caused by the function extraction method, greatly reducing the training time and function space. The storage and security performance of the shared IoT system to encrypt English text will increase with the number of model iterations. Therefore, this work designs the English translation teaching-oriented shared IoT system using a DL-based TC. The finding plays an important role in subsequent English translation and enriching the theory of IoT.

1. Introduction

China's cultural, political, and ideological exchanges extend globally with domestic development and further opening up. Thus, as the demand for composite multilinguistic talents spikes, the importance of college English translation teaching also increases. Reforming college English translation and improving students' comprehensive English quality are keys to translation teaching. This has attracted all stakeholders in English education and international communication [1].

Amin et al. [2] proposed the deep learning- (DL-) based Word2Vec model to convert words into vectors, didimensionalized the high-dimensional sparse vectors, and judged the semantic similarity between words through the word distances. Then, the Doc2vec mode was used to represent sentence vectors and text vectors. Meanwhile, the similarity between sentences/texts was judged through the sentence/text vector distance. Although Word2Vec could lend well to lexical analysis, it effectively failed to use global lexical

information. Shen et al. [3] designed the glove method by learning word vectors based on global vocabulary information, and the representation of word vectors has indeed been improved. Wang et al. [4] constructed a FastText structure similar to Word2Vec, only with different tasks. FastText accelerated the training speed while maintaining high accuracy. It could be as accurate as the DP classifier by using n-grams to narrow the linear and DL models' accuracy gap. Kim et al. [5] utilized a hierarchical attention model to classify long text and represent text features. Compared with other traditional models, the hierarchical attention model performed better in document classification tasks. Zeng et al. [6] applied DL to the text classification (TC) model and greatly improved the classification effect. These research findings had strong research significance in the TC field. With the emergence of artificial intelligence (AI) and boosting series algorithm models, many researchers used the boosting algorithm for TC and achieved good results. Mahdi et al. [7] applied the DL-based TC algorithm to English translation teaching and improved students' classroom

enthusiasm and understanding of English words and grammar. Following the above literature review, this work tries to distinguish itself by further exploring the possible combination of DL-based TC algorithm with the Internet of Things (IoT), Word2Vec, Doc2vec, and attention mechanism (AM). Meanwhile, this work starts from the overall situation in the English translation types without a specific division of poetry, novels, and narrative. Thus, the findings have a certain universality.

DL, TC, and IoT technology are being integrated into English translation teaching in this context. Then, it classifies the English text corpus of Northwestern Polytechnic University using the TC method. The translation results are analyzed by considering the IoT system storage and safety factor. The purpose is to provide effective methods for English translation teaching in the future. Figure 1 displays the theoretical framework.

2. Basic Concept and Technical Route

2.1. TC. Text representation, the primary task of TC, converts text information into a computer-understandable form and then uses various algorithms to complete natural language processing (NLP). TC divides a text into multiple categories according to the subject and content of a given document. In the late 1950s, word frequency statistics, probability models, and factor decomposition algorithms promoted the rapid development of TC technology. Since then, many researchers have carried out extensive research in TC [8].

Latent Dirichlet Allocation (LDA), an unsupervised machine learning (ML) technology, can extract topic information hidden in large-scale document sets or corpora. LDA introduces word exchange by representing each document as a word frequency vector. Then, hard-to-model texts are digitalized [9] to reduce complexity by ignoring the relationship and order among words, thus improving the model. The formation process of LDA is as follows.

(1) For each document, a topic is extracted from the topic distribution

(2) Then, a word is extracted from the word distribution of the topic

(3) The process iterates until each word in the document is traversed

In English translation, the keywords are the text features reflecting the content. The keywords and word frequency differ from document to document. Thus, they are good metrics for distinguishing documents. The words can be extracted through word segmentation. So far, multiple mature English thesauruses are available for word segmentation, such as the data thesaurus [10].

The research on Chinese TC technology starts late and is thus relatively shallow. In developed countries where TC technology has seen much early development, relevant algorithms are mainly used in English TC. Researchers have improved the existing algorithms for Chinese TC. Some

have found mature applications, which have accelerated the development of TC technology in China. Although several ML models have been used in TC, the processing method is relatively simple [11]. It is not robust enough for complex practical problems, such as multicategory medical TC and unevenly distributed texts. Moreover, shallow ML models and integrated classifiers have limited the generalization ability. Fortunately, DL can lend well to TC tasks by scientifically organizing and managing big data. It can also mitigate text information overload. Besides, the DL model has unique advantages in feature extraction and semantic mining. Therefore, the DL-based TC model is the general trend and worth further exploration [12].

2.2. DL. DL, a research hotspot in ML, is a general term for learning methods based on the artificial neural network (ANN). Proposed in the 1940s, DL can simulate the cognitive mechanism of the human brain [13]. Figure 2 is the DL theoretical model.

DL algorithm observes the neural network (NN)'s error function gradient descent and corrects the weight and deviation according to the gradient descent. The specific calculation is given in Equation (1).

$$A_{a+1} = A_a - c_a d_a. \tag{1}$$

In (1), A_{a+1} represents the weight and deviation of the neuron layer. A_a is the weight and deviation after iterative calculation. c_a represents the learning rate of the neural network, and d_a is the gradient of the error function.

Equation Equation (2) is the specific calculation method of the space size of the convolution kernel after the cavity.

$$B = (n - 1) * b + 1. \tag{2}$$

In (2), B represents the size of the convolution kernel space, n is the size of the original convolution kernel, and b is the void rate. Equation Equation (3) calculates the image size after convolution.

$$C = \frac{(l - K + 2P)}{T + 1} \tag{3}$$

In (3), C represents the image size after convolution, and l and K are input and output image sizes, respectively. P is the filling size of the image, and T means the step size.

Nonnegative matrix factorization (NMF), another DL method, can regenerate output as similar as to the original matrix by finding two or more nonnegative matrices to reduce the nonlinear data dimension. Equation (4) calculates the norm cost function based on matrix difference.

$$\min_{P,W,Q} \left\| Y - LWQ^T \right\|_F^2. \tag{4}$$

In (4), Q^T is the transpose of matrix Q. $\|.\|_F$ represents the norm of matrix difference. L, W, Q, and Y represent different matrices. Common DL models are as follows.
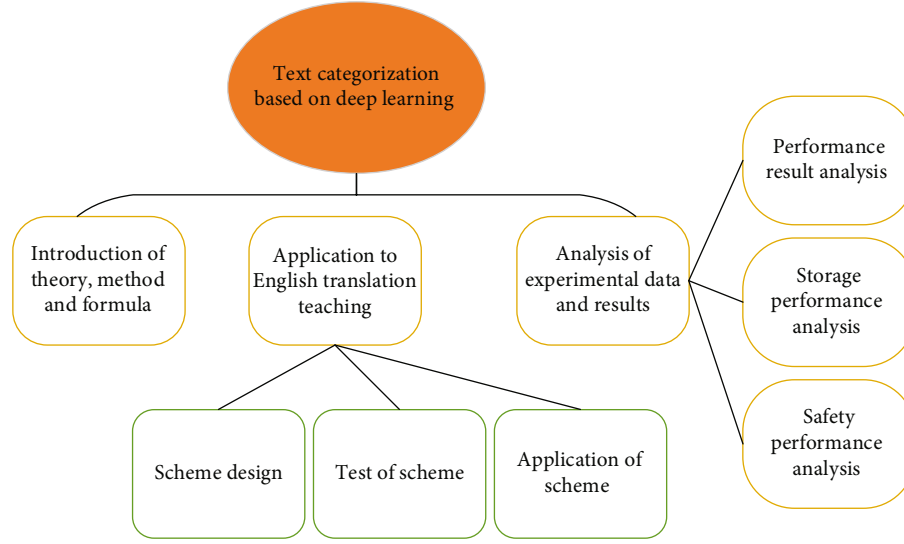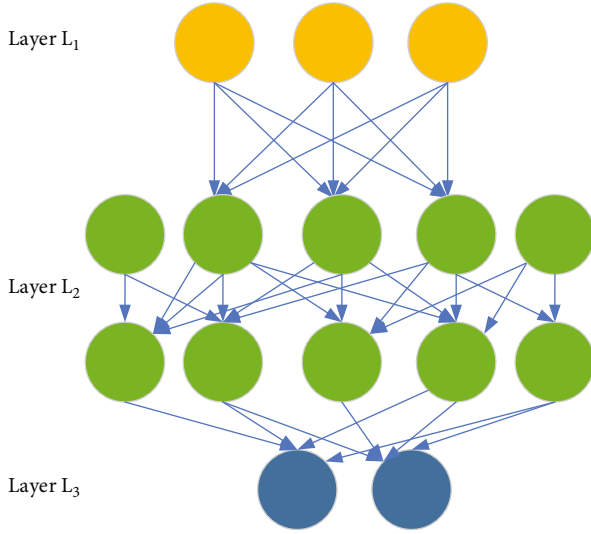
FIGURE 1: Theoretical research structure.

FIGURE 2: DL model.

(1) Convolutional neural networks (CNN): CNN includes convolution calculation and calculates the feedforward neural network with depth structure. It is mainly divided into the input, convolution, pooling, and fully connected layers. It is one of the representative DL algorithms of DL. Figure 3 shows the CNN model

Equation Equations (5) and (6) give the specific calculation.

$$R_t = \mathfrak{I}\left(B_t[b_{t-1}, x_t] + P_f\right), \tag{5}$$

$$Y_t = \mathfrak{I}(A_I \cdot [b_{t-1}, x_t] + P_I). \tag{6}$$

R_t represents the forget gate, $\mathfrak{I}$ is the hidden layer neuron, and B_t is the model output. b_{t-1} represents the hidden layer information at time $t - 1$, x_t is the input at time t, and Y_t is the input gate. P_f and P_I represent the activation functions of the forget gate and input gate, respectively.

(2) The autoencoder (AE): NN based on multilayer neurons includes autoencoder and sparse coding, which have recently attracted extensive attention. Figure 4 is an AE

(3) Deep belief networks (DBN): it is pretrained in the way of a multilayer AE and then combined with identification information to further optimize the network weight. The model can lend to both unsupervised and supervised learning [14]. Figure 5 reveals the DBN model

2.3. IoT Technology. IoT realizes the omniconnection of in and between things and people and between the real and virtual worlds. Data acquisition and object perception are key to IoT, supported by advanced technologies and communication networks. Finally, the collected data can be processed and analyzed to help humans make decisions. The technical means used to realize IoT is called IoT technology [15]. It has many applications involving optical, mechanical, electrical, information, communication, materials, and chemistry fields. Usually, an IoT system is based on the Internet, using sensors, radio frequency identification (RFID), Global Positioning System (GPS), and communication technology. It can provide multiple communication channels [16].

Professor Asshton from the United States first proposed the concept of connecting all sensors to the Internet, making communication between machines possible, and connecting items to the Internet according to the pre-defined protocol. Data are collected on the IoT with RFID and sensor technology for intelligent identification, positioning, and monitoring [17]. Then, IoT combines the physical, virtual, and
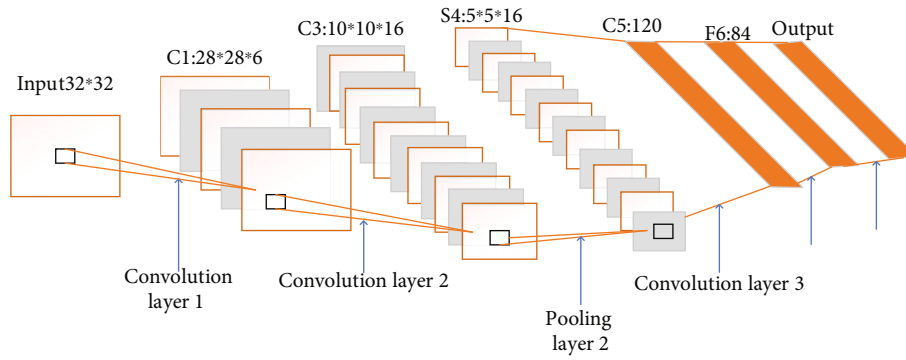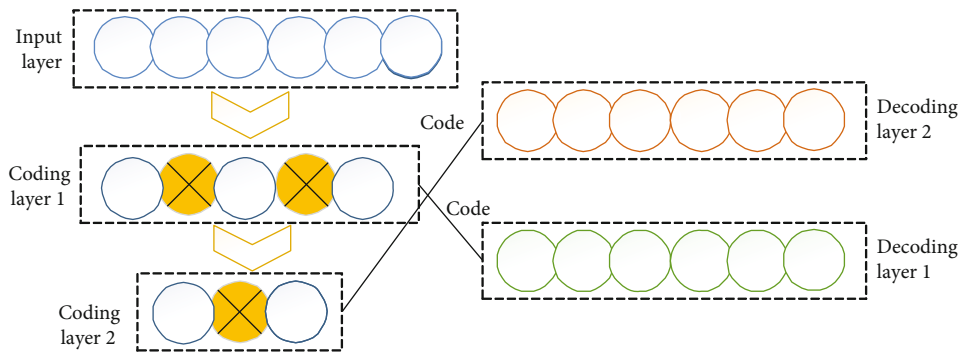
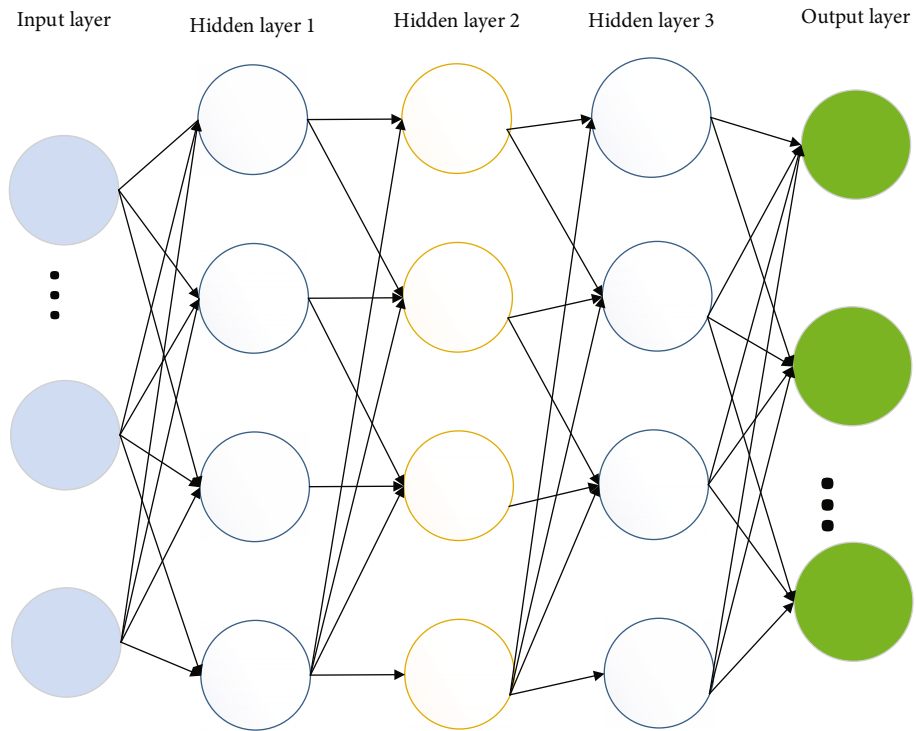Figure 3: CNN model.

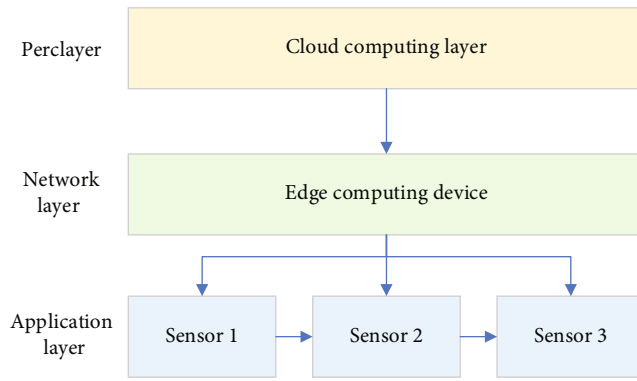Figure 4: AE model.

Figure 5: DBN model.

FIGURE 6: Shared IoT architecture.

digital worlds with human society closely through simple and timely interactions. IoT technology aims to bridge the gap between enterprise big data and individual-user data in the trading platforms [18]. Large enterprises often recruit technical personnel to deal with data transaction business. Most enterprise-level data are not patented, attracting many data peddlers to use data as a resource effectively. Besides, the third-party data trading platform is convenient for individual operators to grasp data [19] comprehensively, even though risks are inevitable, such as data leakage.

2.4. Current Situation of English Translation Teaching. Cultivating students' comprehensive English quality has been put on the agenda in higher education. Traditionally, exam-oriented education focuses on listening, reading, writing, and academic scores while ignoring students' practical skills [20]. In particular, translation is a science with a strong local flavor. In other words, second language learners tend to translate the source language into the target language word-to-word, with poor wording and sentence structures, leading to some misunderstanding. Thus, proper context-based translation is in high demand, especially as China encourages more and more international companies into domestic markets. English translation ability is becoming a basic skill for social talents, just as driving and computer proficiencies. This is where College English must improve its teaching quality in translation, help cultivate multilinguistic talents, and promote China's revitalization. There is a need for higher institutions to offer various English translation practices and introduce English translation methods into classroom teaching. English translation is the main manifestation of students' comprehensive English proficiency and evaluation of their English practical abilities [21]. Most importantly, more opportunities should be provided for students to practice translation outside the classroom, cultivate their interest in learning, and improve their learning enthusiasm.

3. Scheme and Test of IoT System in English Translation Teaching

3.1. Shared IoT Design. As an emerging trading market in recent years, the data trading platform is in the early stage of development. IoT-based data trading platform mostly adopts centralized architecture, completing data storage and transmission through a central control [22], which has

some problems. For example, centralized services cannot catch up with the dynamic needs of data multiplication. Thus, shared IoT provides a new solution and is chosen here for studying the big data-based English translation. Share IoT has a unique structure, as explained below [23].

The perception layer acquires the front-end data using RFID to obtain electronic tags and using Beidou to obtain longitude and latitude, and environmental monitoring temperature, and humidity sensors.

Network layer servers as the background server that transmits information through telecommunication networks and the Internet. It can transmit and process the information from the perception layer. Its key technology is long-distance, high-fidelity transmission, and data processing.

The application layer processes the information from the front-end perception layer and realizes specific applications, such as autonomous driving (AD), environmental monitoring, and health management. It analyzes and processes information, makes correct decisions, or controls behavior to achieve intelligent management and service.

IoT devices need to connect to the Internet to realize remote control. Mobile phones or computers can issue control commands through the Internet. Communication is divided into uplink and downlink, which are the two directions of communication. Transmitting the data on the device end to the cloud is called uplink, and downlink is to send control commands to the user terminal devices, such as the mobile phone or computer. Figure 6 displays the architecture of the Shared IoT.

This section selects the English text corpus of Northwestern Polytechnic University for analysis based on the text encryption algorithm. English text corpora include several categories: agriculture, art, communication, computer, economy, education, electronics, energy, environment, and history. Computers and wireless networks are used as the hardware environment. These data are collected by a computer and then preprocessed, mainly through text retrieval. Finally, according to the needs of this experiment, the text is classified and sorted out. Some categories are relatively few and unrepresentative. Thus,100 representative text categories are selected for the experiment. Figure 7 gives the flowchart of the text encryption algorithm.

3.2. IoT in English Translation Teaching. With the popularization of the 5G network, IoT sees broader applications. In particular, there are three main application forms to English translation teaching:

(1) Basic application: text monitoring: IoT technology can give an alarm in selecting English translation texts once the text number exceeds the preset threshold and help administrators control texts through remote monitoring

(2) Intermediate application: text statistics: IoT systems can analyze the collected text data from different dimensions and types and visualize them through a chart on the screen. As such, administrators can quickly and intuitively understand the operation status of the whole IoT device
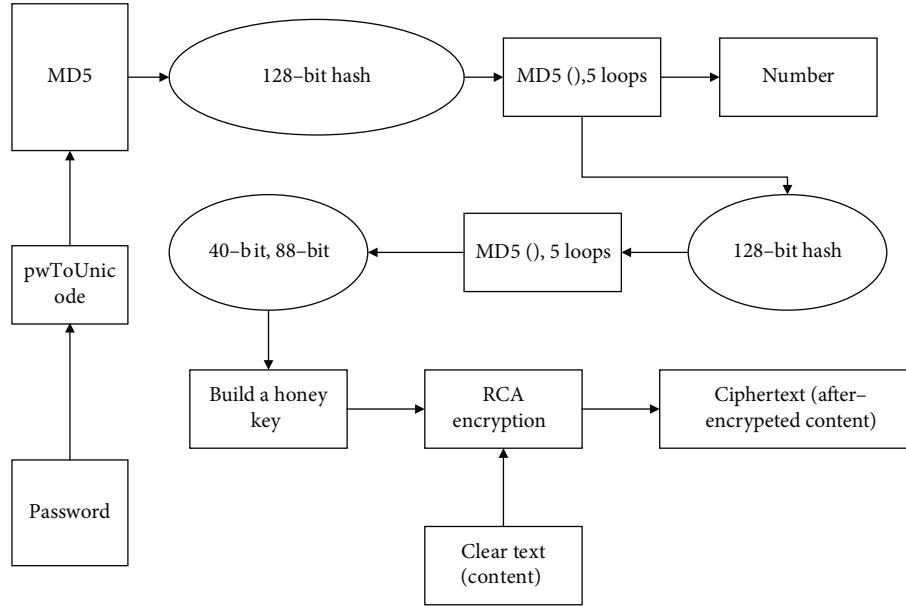
FIGURE 7: Flow chart of text encryption algorithm.

TABLE 1: Experimental data set.

Training set		Test set	
Type	Number	Type	Number
Agriculture	500	Education	500
Art	500	Electronics	500
Signal communication	500	Energy	500
Computer	500	Environment	500
Economics	500	History	500

TABLE 2: Required experimental environment.

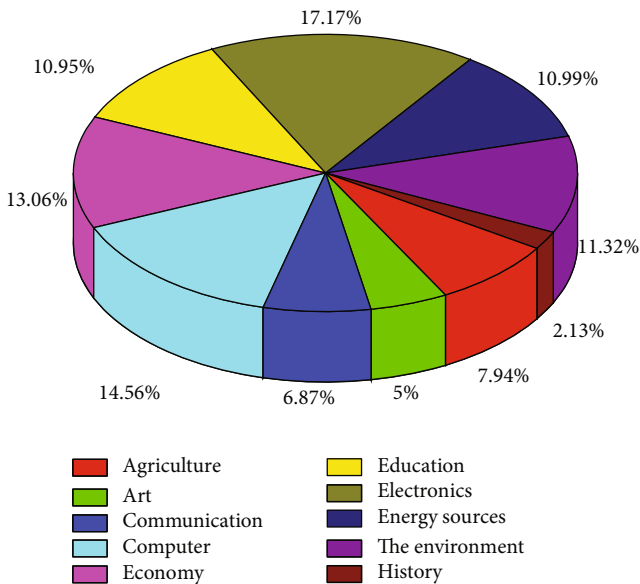| Project	Model
Outward expansion flash	512 K*16
Led liquid crystal display	\
JTAG simulator interface	\
TMS320VC5402@100MHzDSP Main controller	\
Outward expansion SRAM	256 K*16

FIGURE 8: Test results of English translation text.

(3) Advanced applications: data mining: IoT can mine useful things from the database. For example, according to students' information retrieval history, teachers can classify, track, and analyze the English text during translation teaching and provide personalized teaching schemes.

3.3. Verifying the Shared IoT Design. The public's archive utilization needs are very much personalized and diversified. Intelligent and accurate file classification is based on professional knowledge and creativity and employs virtual space and knowledge maps. The archives are mapped from the physical space to the network space based on the Internet and cloud platforms.

In particular, the card in file classification uses an encryption mechanism to encrypt teaching documents. The single key-based encryption has a deadly defect that is prone to attack. The attack is based on password guessing. Each guess is compared with the real user password to calculate the initialization vector. Then, the initialization vector is checked for correctness: the guess and attack do not stop until the vector is correct, and the correct password is returned.
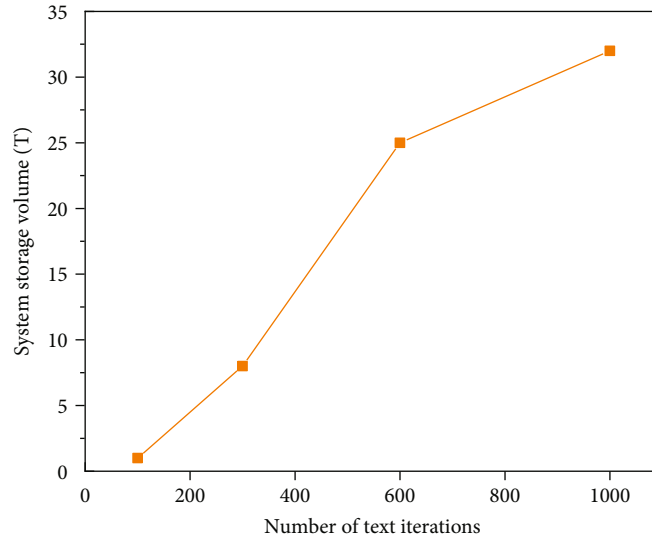
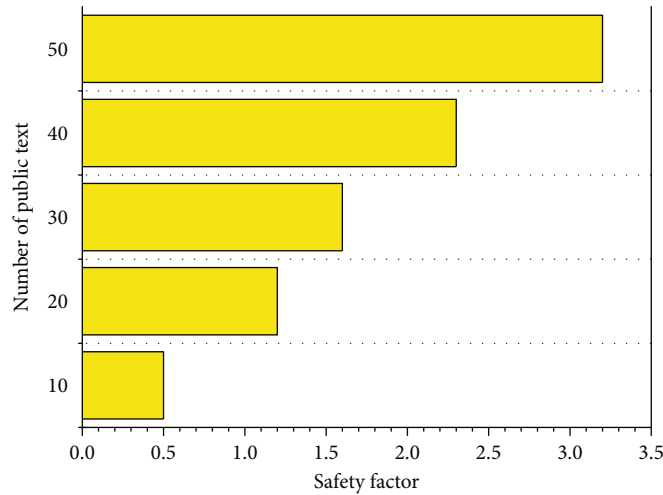FIGURE 9: Storage performance analysis of shared IoT.

FIGURE 10: Analysis of the safety factor of shared IoT maintenance.

4. Experimental Results and Analysis

4.1. Experimental Dataset and Collection. This section takes the English text corpus of Northwestern Polytechnic University as the research object. It uses a text encryption algorithm to analyze the data. The specific experimental data set is shown in Table 1.

Firstly, the IoT system is used to classify the English text corpus of Northwestern Polytechnic University, delete the wrong words, sentences, and grammar, and then train the DL model. Finally, the TC method is introduced to classify the text. Figure 8 gives the details.

In Figure 8, English texts of agricultural, artistic, social, computer, economic, education, electric power, energy, environment, and historical type account for 7.94%, 5%, 6.87%, 14.56%, 13.06%, 10.95%, 17.17%, 10.99%, 11.32%, and 2.13%, respectively. Obviously, the proposed TC method overcomes the performance loss caused by the function extraction method, avoids the problem that the function extraction method does not consider the semantic relationship between words, greatly reduces the function space, and reduces the training time. Experiments show that the classification effect is basically unchanged, and the TC method is better than the function extraction method.

4.2. Experimental Environment and Storage Analysis of Shared IoT System. The specific experimental environment is shown in Table 2.

Figure 9 analyzes the storage performance of the shared IoT.

In Figure 9, the storage analysis curve of the IoT shows a rising trend as a whole. The shared IoT system consumes 1 T, 8 T, 25 T, and 32 T storage capacity to encrypt the text through 100, 300, 600, and 1,000 iterations. Thus, the storage of the shared IoT system increases with the number of iterations to encrypt a text.
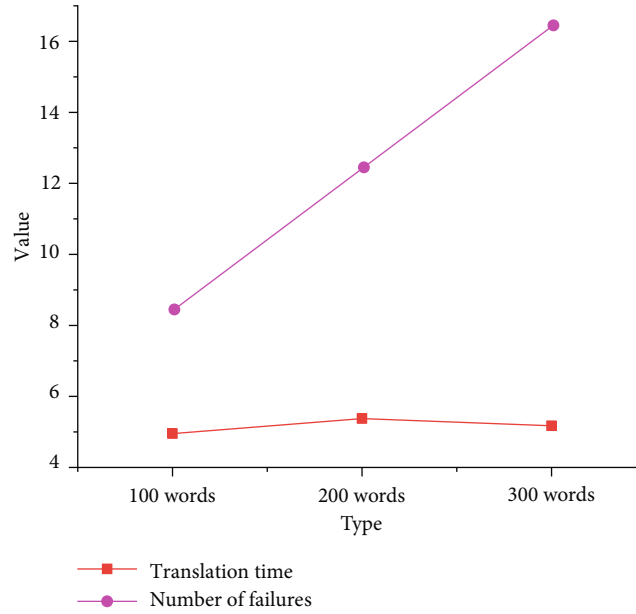
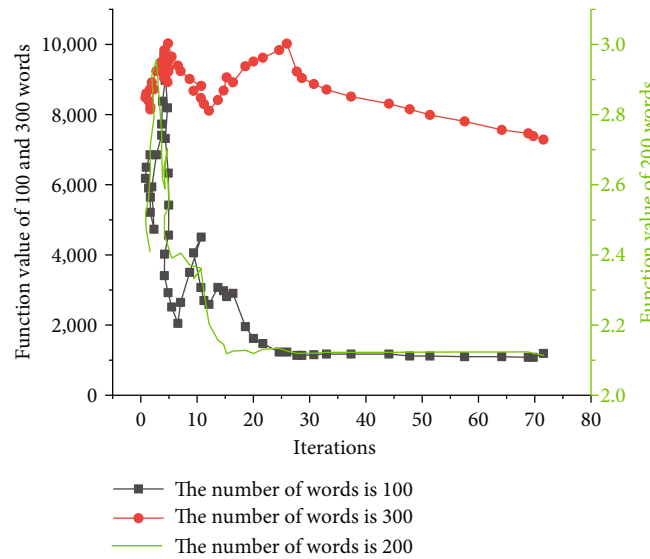FIGURE 11: Translation time and failure times under different word numbers.

FIGURE 12: Iterations under different word numbers.

4.3. Analysis of Security Factor of Shared IoT Maintenance. Figure 10 analyzes the safety factor of shared IoT maintenance.

The security factor of the shared IoT system is 0.5, 1.6, and 3.0 to encrypt 10, 30, and 50 texts. Thus, the shared IoT system's security factor increases with the number of encrypted texts.

In order to analyze the performance of the TC more accurately, the simulated annealing algorithm (SAA) is now introduced. Then, 100, 200, and 300 English words are selected as the research data. The specific experimental results are shown in Figure 11.

According to Figure 11, when the number of translated words is 100, 200, and 300, the translation time and failure times are 4.953 seconds and 8, 5.375 seconds and 12 and

5.172 seconds and 16, respectively. Thus, the translation time does not significantly increase with the increase in word number, maintained at around 5 seconds on average. At the same time, translation failures increase with the number of translated words.

The iteration times of SAA under a different number of translated words are shown in Figure 12.

Apparently, SAA's translation speed declines less obviously with the increase of translated words. The algorithm iteration is stable no matter how many words. Specifically, when translating 100, 200, and 300 words, SAA reaches the optimal solution at around the 30th iteration. Thus, the number of translated words has a small impact on the algorithm iterations.

5. Discussion

English translation is one of the critical contents of English teaching. Here, 100 text categories are selected from the English text corpus of Northwestern Polytechnical University and are analyzed by integrating DL, TC, and IoT. The result shows that the TC method can reduce the training time and speed up model convergence. The storage and security performance of IoT encrypted text will increase with the number of English texts. Cherif et al. [24] proposed a TC model based on semantic enhancement and feature fusion. The model obtained the enhanced semantic features of the CNN model through the attention layer and then fused them with the local features extracted by the convolution layer. Then, experiments were conducted to show its effectiveness. Lee et al. [25] constructed a new word vector model for Chinese text based on the structure of Word2Vec. They found that the partial radicals of Chinese characters contain certain semantic information and introduced the partial radical information into the text representation. Experiments showed that the model had a good classification effect for classifying Chinese news headlines. Luo [26] proposed a DL CNN model to enrich the text generalization and memory information, giving the word vectors richer stylistic features. Compared with the theories and methods proposed by the three scholars, the proposed DL-based TC is simpler and easier to use in practice. The TC algorithm has been applied in many fields, providing rich references. Additionally, IoT technology is one of the most key technologies. The continuous promotion of the 5G network greatly facilitates IoT development. Thus, applying IoT in English translation teaching is the development trend of the times. Finally, there are not many theories of applying DL and TC to the field of the English translation, where this work bridges the gap.

6. Conclusion

The purpose is to design shared IoT to classify, translate, and search English texts efficiently, accurately, and quickly. This work first selects 100 text categories from the English text corpus of Northwestern Polytechnic University and analyzes them using the DL-based TC. Then, it describes the storage and security performance of the IoT system. The main conclusions are as follows. The storage capacity of the shared IoT system increases with the encrypted text iterations, and the security coefficient storage of the shared IoT system also increases with the number of encrypted text. The result proves that the designed IoT system is safe and effective. Additionally, using the proposed DL-based ANN method in classifying the English corpus can reduce the model training time.

The deficiency is summarized. Certain limitations exist in the collected text types. Research on other fields is lacking, such as Chinese TC. The future research direction is to improve the retrieval speed of massive amounts of TC encryption in the shared IoT system. Then, the shared IoT system should be optimized further, ensure against data duplication and resell transactions of the third-party platform, and reduce user risk.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Consent

Informed consent was obtained from all individual participants included in the study.

Conflicts of Interest

All authors declare that they have no conflict of interest.

Acknowledgments

References

[1] D. Kwesiga, C. Tawiah, M. A. Imam et al., "Barriers and enablers to reporting pregnancy and adverse pregnancy outcomes in population-based surveys: EN-INDEPTH study," *Population Health Metrics*, vol. 19, no. S1, pp. 1–14, 2021.

[2] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. S. Hossain, "Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion," *Future Generation Computer Systems*, vol. 101, no. 4, pp. 542–554, 2020.

[3] C. Shen, G. Song, X. Zhu et al., "An in-depth study of heteroatom boosted anode for potassium-ion batteries," *Nano Energy*, vol. 89, no. 78, article 105294, 2020.

[4] S. Wang, B. Kang, J. Ma et al., "A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19)," *European Radiology*, vol. 12, no. 1, pp. 12–15, 2021.

[5] D. W. Kim, G. Lee, S. Y. Kim et al., "Deep learning–based algorithm to detect primary hepatic malignancy in multiphase CT of patients at high risk for HCC," *European Radiology*, vol. 3, no. 18, pp. 1–11, 2021.

[6] X. Zeng, Z. Jiang, W. Luo et al., "Efficient and accurate identification of ear diseases using an ensemble deep learning model," *Scientific Reports*, vol. 11, no. 1, p. 10839, 2021.

[7] F. P. Mahdi, K. Motoki, and S. Kobashi, "Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs," *Scientific Reports*, vol. 10, no. 1, pp. 22–30, 2020.

[8] A. Dhar, H. Mukherjee, N. S. Dash, and K. Roy, "Text categorization: past and present," *Artificial Intelligence Review*, vol. 54, no. 4, pp. 3007–3054, 2021.

[9] L. Xia, R. Zhao, Y. Wu, and X. Tong, "Intentional control of type I error over unconscious data distortion: a Neyman-Pearson approach to text classification," *Journal of the American Statistical Association*, vol. 116, no. 533, pp. 68–81, 2021.

[10] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep learning–based text classification," *ACM Computing Surveys (CSUR)*, vol. 1, no. 1, p. 42, 2020.

[11] C. Tsigkanos, M. Garriga, L. Baresi, and C. Ghezzi, "Cloud deployment tradeoffs for the analysis of spatially distributed Internet of Things Systems," *ACM Transactions on Internet Technology (TOIT)*, vol. 20, no. 2, pp. 1–23, 2020.

[12] Y. Liang, H. Li, B. Guo et al., "Fusion of heterogeneous attention mechanisms in multi-view convolutional neural network for text classification," *Information Sciences*, vol. 548, pp. 295–312, 2021.

[13] C. Chang and M. Masterson, "Using word order in political text classification with long short-term memory models," *Political Analysis*, vol. 28, no. 3, pp. 395–411, 2019.

[14] J. Xu and Q. Du, "Adversarial attacks on text classification models using layer ise relevance propagation," *International Journal of Intelligent Systems*, vol. 35, no. 9, pp. 1397–1415, 2020.

[15] M. A. Ibrahim, M. U. Khan, F. Mehmood, M. N. Asim, and W. Mahmood, "GHS-NET a generic hybridized shallow neural network for multi-label biomedical text classification," *Journal of Biomedical Informatics*, vol. 116, no. 1, article 103699, 2021.

[16] F. Kanavati, G. Toyokawa, S. Momosaki et al., "Weakly-supervised learning for lung carcinoma classification using deep learning," *Scientific Reports*, vol. 10, no. 1, pp. 26–30, 2020.

[17] N. Koroniotis, N. Moustafa, and E. Sitnikova, "A new network forensic framework based on deep learning for Internet of Things networks: a particle deep framework," *Future Generation Computer Systems*, vol. 110, no. 8, pp. 91–106, 2020.

[18] P. Lam, A. Zhu, L. Salminen, S. Thomopoulos, N. Jahanshad, and P. Thompson, "Comparison of deep learning methods for brain age prediction," *Biological Psychiatry*, vol. 87, no. 9, pp. 374–S375, 2020.

[19] X. Kong, Y. Xu, Z. Jiao, D. Dong, X. Yuan, and S. Li, "Fault location technology for power system based on information about the power Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6682–6692, 2020.

[20] A. Smerdin, G. Ermachkov, V. Nezevak, O. Sidorov, and A. Golubkov, "Use of sensor networking technology to build a power transformer monitoring system," *E3S Web of Conferences*, vol. 224, no. 4, article 02021, 2020.

[21] S. Moon, M. Y. Kim, and D. Iacobucci, "Content analysis of fake consumer reviews by survey-based text categorization," *International Journal of Research in Marketing*, vol. 38, no. 2, pp. 343–364, 2021.

[22] L. Enamoto, L. Weigang, and G. P. R. Filho, "Generic framework for multilingual short text categorization using convolutional neural network," *Multimedia Tools and Applications*, vol. 80, no. 9, pp. 13475–13490, 2021.

[23] T. Wang, Y. Cai, H. Leung, R. Y. K. Lau, H. Xie, and Q. Li, "On entropy-based term weighting schemes for text categorization," *Knowledge and Information Systems*, vol. 63, no. 9, pp. 2313–2346, 2021.

[24] W. Cherif, A. Madani, and M. Kissi, "Supervised classification by thresholds: application to automated text categorization and opinion mining," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 4, 2022.

[25] Y. H. Lee, P. J. H. Hu, W. J. Tsao, and L. Li, "Use of a domain-specific ontology to support automated document categorization at the concept level: method development and evaluation," *Expert Systems with Applications*, vol. 174, no. 8, article 114681, 2021.

[26] X. Luo, "Efficient english text classification using selected machine learning techniques," *Alexandria Engineering Journal*, vol. 60, no. 3, pp. 3401–3409, 2021.