Analysis of English Translation of Corpus Based on Blockchain

Aiping Zhang, Changsha Social Work College, China
Xiaowei Zhu, Wuhan Polytechnic, China*

ABSTRACT

Blockchain technology can create a shared platform for English translation and reserve a large number of practical corpus resources, thus improving the quality of machine translation. This paper first introduces the research status of foreign language corpus and blockchain English translation in China. Then it introduces the basic principles of BPNN and particle swarm optimization and constructs the PSO-BP model. Experiments show that the prediction accuracy of BPNN optimized by particle swarm optimization algorithm is greatly improved, the convergence speed is faster, and it will not fall into the local optimal trap. Finally, this paper proposes the implementation path of blockchain in corpus translation application: (1) build "blockchain+ AI" English translation corpus and (2) improve the machine English translation software of the "blockchain+ AI" English translation training platform.

KEYWORDS

Blockchain Technology, Corpus, Deep Learning, Translation

INTRODUCTION

English has become an essential communication tool in international conferences, forums, and academic seminars, in international business negotiations and international business contract texts, and in the work of foreign-funded enterprises or joint ventures (Lyons&Urry, 2005). With the rapid development of world economic integration, China has been more widely integrated into the international community, and exchanges with countries worldwide in political, economic, cultural, and other fields have become increasingly frequent. Industries related to English majors, such as foreign trade, diplomacy, customs, tourism, management, and other foreign-related departments, have obtained unprecedented development opportunities, increasing China's demand for English talent. English translation talents also take on new characteristics (Jiang, 2020). They need to have multiple skills such as multilingual communication, cross-cultural communication, international business management, cross-border investment cooperation, and cross-border flexible innovation to adapt to the development of the new era (Wei et al., 2022).

In the 1990s, just as the philosophy circle experienced the linguistic turn, the translation circle also experienced the cultural turn, which inadvertently brought a new perspective to translation

DOI: 10.4018/IJWLTT.332767

*Corresponding Author

studies. Translation research has gradually got rid of the shackles of "prescriptive research," and scholars in the translation field have increasingly recognized and applied "descriptive research" (Wan et al., 2021). In descriptive research, the translated text is no longer regarded as a derivative of the original text and gradually moves from the edge of translation and culture to the center. The multi-system theory holds that the translated text is an integral part of the target language culture and an independent subsystem interacting with the target language culture. This view enriches and expands the field of translation research and provides methodological support and a foothold for studying translated texts independent of the original text. It is against this background that corpus translation studies are produced (Baker, 2019). Corpus translation research has two pillars: corpus linguistics and descriptive translation research. Several scholars use the corpus to discuss translation norms and commonalities, including comparable corpus, bilingual parallel corpus, and multilingual corpus (Du et al., 2017). With the deepening of corpus translation research, researchers have gradually found that corpus is also vital in translator training. This emphasis puts corpus research into the category of applied translation research, enriches the original research dimension, and expands the application of corpus in translation teaching research and practice. Using monolingual corpus alone can help translators improve their translation skills and make the translation closer to the natural target language (Bowker, 2001). At the same time, the researchers found that the corpus can also be used as a standard to evaluate students' and translators' translated works (Lin et al., 2020). Regarding training translators based on a corpus, some researchers have tried to guide students to build a corpus suitable for their teaching field (Lin et al., 2020). In addition, researchers have found that corpora can be used as a standard to evaluate students' and translators' translation (Peterlin, 2010). As an evaluation tool, a corpus is of great value in evaluating the quality of students' translated works (Duro Moreno, 2020). First, teachers and students can use the corpus to verify the authenticity of their language sense. Secondly, teachers can use the corpus to verify and explain why one translated text is better and reflects the target language more truthfully. Corpus-based teaching research has become the main trend of empirical research in teaching research.

Blockchain is a distributed structured database composed of pointers and blocks. It records and stores data in chronological order, uses blockchain to verify data, uses consensus algorithm to update data, uses smart contracts to program and operate data, and uses cryptography to protect data access (Jung & Agulto, 2019). The blockchain adopts distributed accounting and storage, and each block operates independently, so there is no central system or organization, and it has the characteristics of decentralization (Green et al., 2015). With the development of artificial intelligence technology, the quality of machine translation has improved steadily in the past few years. China's artificial intelligence (AI) translation has entered a new era by implementing the Belt and Road Initiative. Machine translation models and corpus construction are two important factors that restrict the improvement of AI translation technology. In the field of English translation, AI translation technology can improve the service in this field. AI translation technology makes it easier for individuals to process massive data and perform related tasks. Despite the rapid development of the global digital economy, multinational enterprises still encounter translation problems. Blockchain and AI technologies are still in the early stages of research, but it is conceivable to achieve document decentralization in international communication using these technologies (Brynjolfsson et al., 2019).

Since the idea of combining blockchain and artificial intelligence technology was put forward in 2017, many industries have begun to explore how to combine the two to promote social development. The application scenario of blockchain technology combined with artificial intelligence technology in the field of translation will become a new exploration direction of the big data economy.

This paper first introduces the research status of foreign language corpus and blockchain English translation in China. Then it introduces the basic principles of BPNN and PSO and constructs the PSO-BP model. Finally, this paper proposes the implementation path of blockchain in corpus translation applications.

RELATED WORK

English translation and corpus linguistics are closely related and complement each other. On the one hand, the research of English translation increasingly uses the methods and data of corpus linguistics; on the other hand, various comparative studies are often carried out in corpus linguistics research (Busari et al., 2019). In recent years, corpus-based language comparative research has shown an upward trend. Corpus data has become an essential basis for comparative linguistics, and corpus linguistics methods have become an important research method for English translation learning.

Baker (2019) reviewed the Fifth International Conference on Contrastive Linguistics and summarized the new trends in comparative linguistics research. At this conference, about 60% of the English papers used corpus as the primary basis for comparative analysis. The comparative research based on corpus is the mainstream method of this conference. Two of the themes for the conference involve corpora, Parallel Corpus in Language Comparison, Contrastive Linguistics, and Corpus Analysis and Annotation. The content involves comparative linguistics and corpus translation research. Scholars believe these two disciplines are complementary and their research methods have many similarities (Zanettin, 2013). Furthermore, Baker (2004) summarizes the research progress of corpus-based contrastive linguistics and points out that many theoretical and applied studies have emerged in the past decade. There are parallels and differences between English and Chinese personal pronouns; therefore, the corpus may be used to compare the two.

Regarding frequency of usage, English personal pronouns are more common than Chinese personal pronouns. The frequency of personal pronouns in English and Chinese was studied in two parallel corpora, according to Shlesinger (1998). Results demonstrate that English uses more personal pronouns per 1,000 words than Chinese and that each category's frequency varies widely between the two languages. English third-person pronouns outnumber Chinese counterparts in personal pronoun use, whether in English-Chinese or Chinese-English translations. Compared to Chinese, English second-person and first-person pronouns are less common. Data from Wang (2021) were used to determine the number of third-person pronoun anaphora in both English and Chinese. For example, in English and Chinese, using third-person pronouns is far more common than in Chinese. Wang (2021) also pointed out that compared with English, the personal pronouns in Chinese are more omitted, and the subject, object and attribute are all omitted. Possessive pronouns and subjects are omitted frequently, which is a common phenomenon.

Contrastive language research is integral to corpus linguistics research and has also shown a growing trend in recent years. Interlanguage comparative analysis involves comparing the characteristics of the mother tongue and interlanguage; one of the purposes is to promote foreign language teaching. Translation comparative research can compare the language characteristics of different translation directions, and the research results are helpful for translation teaching and translation practice. The research on translation commonality can involve comparing three language forms in two languages, and the research results can also be applied to translation teaching and translation practice (Peterlin, 2010). Since the early 1990s, with the rapid development of corpus linguistics and the continuous influence of descriptive translation studies, a group of scholars have applied the research results of corpus linguistics to translation studies, creating a new paradigm of translation studies, namely Corpus Translatology (Pan & Qin, 2022). One of the advantages of this new research paradigm is that it draws on the research methods and achievements of corpus linguistics and focuses on empirical research. By obtaining translation data from the corpus and then conducting statistical analysis, the research results are more objective and scientific, providing an advantage over traditional translation research. Lin et al. (2020) first proposed corpus translation studies.

According to its development, blockchain technology has the following characteristics. First, information integration. With the development of the developed information technology foundation and economic globalization in the 21st century as the background, blockchain establishes a vast database through network technology. It integrates fragmented and fragmented information on this basis to form a confidential database system (Inwood & Zappavigna, 2021). Second, the continuity of time. When the blockchain integrates data, it adds time conceptualization according to its time sequence so data can be sorted according to time (Sukman et al., 2022). Third, security. The blockchain uses the principles of cryptography to manage the security and confidentiality of the database and adopts anonymous access to protect the data (Duan & Wei, 2021). The underlying design of blockchain in translation applications includes, for example, the application layer, contract layer, incentive layer, consensus layer, data layer, and network layer. Deng Shangyu (2019) proposed that blockchain is a data storage computing system, pointing out blockchain technology's chain rule and decentralization principle. Pak Lefei (2019) pointed out in his paper that blockchain has smart contracts, consensus mechanisms, cryptography principles, distributed accounting, and other major technologies. At the same time, he believed that blockchain technology would have significant future applications. Tran (2022) mentioned that blockchain development is still in its infancy. At present, it cannot provide enough help in the research of business English development model. However, Sun's research will help future generations to study the relationship between business English and blockchain.

METHOD

Corpus-Based Evaluation System for English-Chinese Translations

This study uses the methods of corpus linguistics and statistics to carry out statistical analysis and comparative study on the translated texts of the English-Chinese translation test of college English and their official manual scoring data. Through empirical methods, the characteristics of the translated text concerned in the manual grading of English-Chinese translation are initially determined, and the evaluation model is automatically constructed to verify its predictive power on the quality of the translated text. It aims to achieve the organic combination of macro fuzzy evaluation and micro accurate evaluation of English-Chinese translation in large-scale language testing.

Before manually annotating the translation corpus of the English-Chinese translation test, the papers related to the translation test evaluation in the national core journals of foreign language teaching and research in the past ten years were retrieved. Since this study aims to reveal the features of the translated text concerned in the manual scoring of translation testing, the pragmatics, objectivity, and functional factors that affect the quality of translations are not considered. The translation evaluation standard generally accepted by translation scholars and language teachers is preliminarily identified as a quantitative evaluation factor for translation that may play a role in the manual scoring of the College English-English-Chinese Translation Test. The correlation statistics between the possible quantitative evaluation factors and the formal manual scoring data of the research corpus were carried out. First, using the corpus linguistics method, according to the quantitative evaluation factors concerned with the manual scoring, the corresponding errors from the translated texts of the college English-English-Chinese translation test were manually marked to form a corpus. In order to ensure the accuracy and consistency of mislabeling, the annotators should standardize the corpus labeling before manual labeling, including the definition of conceptual terms, labeling methods, and the usage of labeling symbols. The obtained labeling data is processed by the weighted combination of Bayesian statistical mode, and the labeling accuracy is close to 95%. Then, use the statistical function of Excel to collect and retrieve the errors in the marked familiar corpus according to different quantitative factor types and input the retrieved results and the formal manual scoring data of the same corpus into SPSS. Finally, through Pearson correlation analysis, quantitative evaluation factors significantly related to the formal manual scoring are extracted; that is, the text surface feature items concerned in the manual scoring reflect the quality of the English-Chinese translation, as shown in Table 1.

Table 1. Corpus-based quantitative evaluation system for English-Chinese translations

Index	Label
Misjudgment of parts of speech	L1
Wrong choice of lexical meaning	L2
Mistranslations of plural nouns	L3
Wrongly written characters	L4
Mistranslations of collocations	L5
Uncontrolled translations in phrases	L6
Improper amplification in phrases	L7
Improper collocations in TL	L8
Incorrect syntactic analysis	L9
Incorrect analysis of tenses	L10
Incorrect analysis of complex sentences	L11
Uncontrolled translation at the syntactic level	L12
Random amplification in the syntactic level	L13
Improper divisions	L14
Incorrect punctuations	L15
Incorrect references to pronouns	L16
Thematic inconsistency	L17
Errors at the lexical level	L18
Errors at the phrasal level	L19
Errors at the textual level	L20

BPNN Based on PSO

The Basic Principle of BPNN

An input layer, an output layer, and one or more hidden layers are typical components of a BPNN, a supervised learning technique based on error backpropagation. The nonlinear mapping may be completed by learning to adjust weights and thresholds across nodes continually, reducing the error between anticipated output and ideal output. Two learning steps are involved in implementing the BP algorithm, which uses gradient descent to converge. The first stage is to use the known samples as the input value of the network. By setting the network structure and learning the connection weights and thresholds, the output of each neuron is calculated backward from the input layer through the hidden layer. The second stage is to calculate the influence of the weights and thresholds between layers on the total error from the output layer to the hidden layer according to the target output and the actual error and then correct the connection weights and thresholds between the layers. Through repeated revisions in two stages, the predicted output of the BPNN is constantly approaching the expected output until the network reaches convergence. The output calculation method of the neuron is shown in formula (1).

$$y_i = f\left(\sum_{j=1}^{n} w_{ij} x_j + \theta_j\right) \tag{1}$$

Analysis and Research of PSO Algorithm

Mathematical model of PSO algorithm: Assuming that the particle coordinates of the i particle in the particle swarm are $X_i = \left(x_{i1},\ x_{i2},\ldots,x_{in}\right)$, the best position is recorded as p_1, and the best-recorded position of all particles is p_2, the moving speed of the particle is $V_i = \left(v_{i1},\ v_{i2},\ldots,v_{id}\right)$. In particle search, the particle's trajectory in space is shown in formula (2) and formula (3) in each iteration.

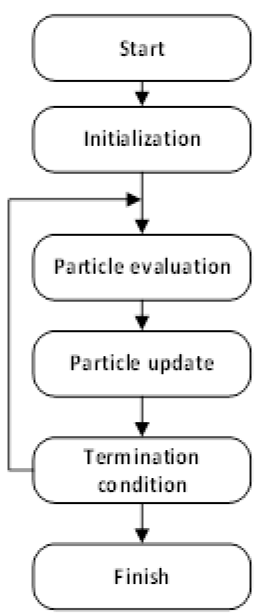
$$v_{id}^{k+1} = v_{id}^{k} + c_1 rand()\left(p_1^k - x_{id}^k\right) + c_2 rand()\left(p_2^k - x_{id}^k\right) \quad (2)$$

$$x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1} \quad (3)$$

where c_1 and c_2 are the acceleration constants, respectively, the purpose is to make the particles move faster to the best position and the best position of all particles; $rand()$ is a random number of [0, 1]; the maximum value of the particle moving speed is $Vmax$.

As soon as the particle's velocity hits the maximum value Vmax, it will stop increasing. It is vital to establish a maximum particle velocity in order to increase the overall search accuracy. An excessively fast particle will miss the optimum value for a given solution space, whereas an excessively slow particle will end up at the local optimal solution. As a result, this configuration is required. The velocity of a particle consists of three elements, the first of which is the particle's starting velocity motion, which characterizes the particle's current motion state. The particle's velocity will stay constant if there is no interference. Particles' cognitive abilities are simulated using a technique that mimics bird cognition. Sharing information is the last step. There will be interactions between particles throughout the optimization process, which might cause the whole particle swarm to develop collectively, reflecting the social traits of particles. The general flow of the PSO algorithm is shown in Figure 1.

Figure 1. General steps of particle swarm optimization



Construction of PSO-BP Model

1) The BPNN is used to estimate the cost module. The number of hidden nodes, hidden layers, the activation function, the learning rate, and the momentum coefficient are some of the BPNN's most essential factors. There are a total of: convergence may take longer with more hidden nodes, but the "training sample" error is reduced. No matter how much you raise the number, the execution time may abruptly increase if the error level rises over a particular point. Formula (4) calculates the number of hidden layer nodes.

$$l = \sqrt{n + m} + a \tag{4}$$

Number of hidden layers: The number of hidden layers has an essential impact on the convergence speed of the network. Usually, one to two layers are ideal and can solve most application problems. One layer is used for training. Activation function: The sigmoid function is used as the activation function, and its calculation formula is shown in formula (5):

$$f\left(x\right) = \frac{1}{1 + e^{-x}} \tag{5}$$

Learning rate and momentum coefficient: The learning rate significantly impacts the network's convergence speed. In general, the learning algorithm will add a momentum coefficient. That is, a certain proportion of the previous weight change is added to weaken the oscillation in the convergence and accelerate the convergence.

2) Randomly generate the initial velocity and position of the particle swarm. Set $X_i = \left(x_{i1},\ x_{i2}, \ldots, x_{in}\right)$ as the initial position of particle i, set $V_i = \left(v_{i1},\ v_{i2}, \ldots, v_{id}\right)$ as the initial velocity of particle i, $P_g = \left\{p_0,\ p_1, \ldots, p_s\right\}$ for all, the best position of the particle is recorded, that is, the local optimum position. The velocity and position vectors of the particle swarm in n-dimensional space are randomly generated at (0, 1).
3) Calculate the fitness of each particle. The fitness of each particle is assessed using the problem-specific objective function. The fitness function value is compared to the best function value in memory, and the particle alters the search speed of the next stage following the best value in memory. The sum of squared errors (SSE) is shown in formula (6):

$$f\left(i\right) = SSE = \sum_{i=1}^{n} \left(T_i - E_i\right)^2 \tag{6}$$

where T_i is the data fitted by the algorithm, and E_i is the original data.

4) Take a reading of the particle's location and velocity vector. If the particle best value is better than the global best value, adjust the global best value in memory and, at the same time, fix the location and velocity of each particle so that the next global search may proceed more quickly and accurately.

$$V_i^{k+1} = WV_i^k + c_1 rand()\left(s_i^{k1} - s_i^k\right) + c_2 rand()\left(s_i^t - s_i^k\right) \tag{7}$$

$$s_i^{k+1} = s_i^k + V_i^{k+1} \tag{8}$$

where c_1 and c_2 are learning factors, also known as acceleration constants, W is inertia factor, $rand()$ is a uniform random number of [0,1], s_i^{k1} represents the k1th dimension of the individual extreme value of the ith variable, s_i^t represents the ith dimension of the global optimal solution, s_i^k represents the kth dimension of the individual extreme value of the ith variable.

The algorithm flow of the PSO-BP evaluation model is shown in Figure 2.

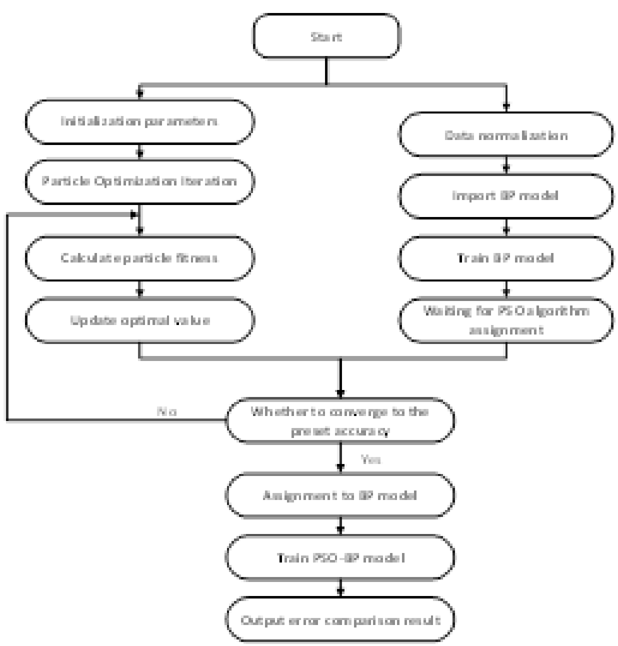
EXPERIMENT AND ANALYSIS

Experimental Data Source and Preprocessing

The reliability of model output outputs is directly related to the quality of the data set used in the neural network method. Excluding data that did not match the translation findings, were ambiguous, or could not be used for feature extraction was done to ensure a diverse sample of 3000 English translations from an Internet company's database. As part of the data augmentation process, such as introducing noise, an additional 6500 testable records are added to the data set to enhance data volume and improve the model's generalizability. The correctness and variety of the English translation feature sample collecting results may be guaranteed to the fullest degree by the data set produced via the approach mentioned earlier. Then, use the following formula to preprocess the data:

$$y_p = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{9}$$

Figure 2. Algorithm flow of the PSO-BP prediction model



Model Parameter Selection

1) Selection of the number of hidden layer nodes: By calculating the formula in the previous chapter, this paper selects the range of the number of hidden layer nodes [5-15] to conduct experiments, and the experimental results obtained are shown in Figure 3.
2) The influence of the number of particles: various parameters must be selected when utilizing the particle swarm optimization approach to train a neural network. It is set to 0.01, and all other parameters are randomly created. Analyze the performance of three distinct swarm sizes, namely 5, 10, and 20, following the convention of choosing particle swarm size by particle swarm algorithm. Therefore, the result is the best possible option at this moment, and the iteration is halted, and the relevant data are recorded, as shown in Table 2.

Model Performance Test

Import the relevant sample and test data into the BP and PSO-BP models to obtain the prediction results. The BPNN converges between the optimal and target errors after iterating for 600 steps, while the PSO-BP model has 400 steps. Accurately converge to the preset target error for a long time, as shown in Figure 4. It is proved that the prediction accuracy of the BPNN is greatly improved after the optimization of the PSO algorithm, the convergence speed is improved, and the local optimization trap is not triggered.

Figure 3. The effect of the number of hidden layer nodes on the model error

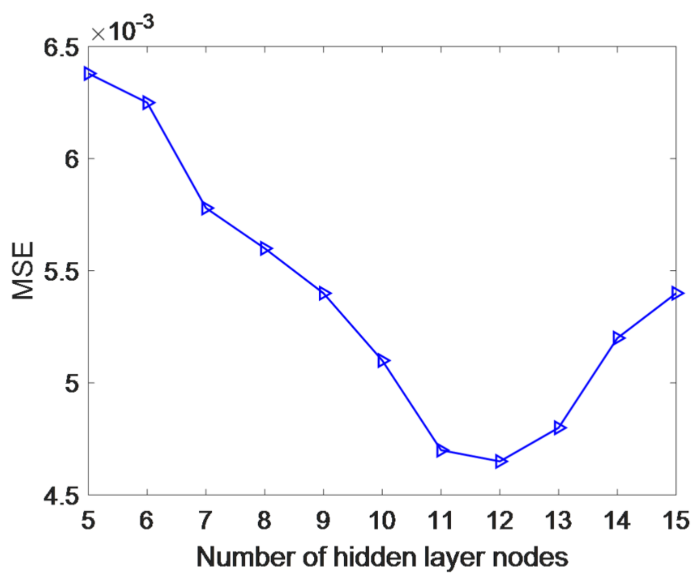
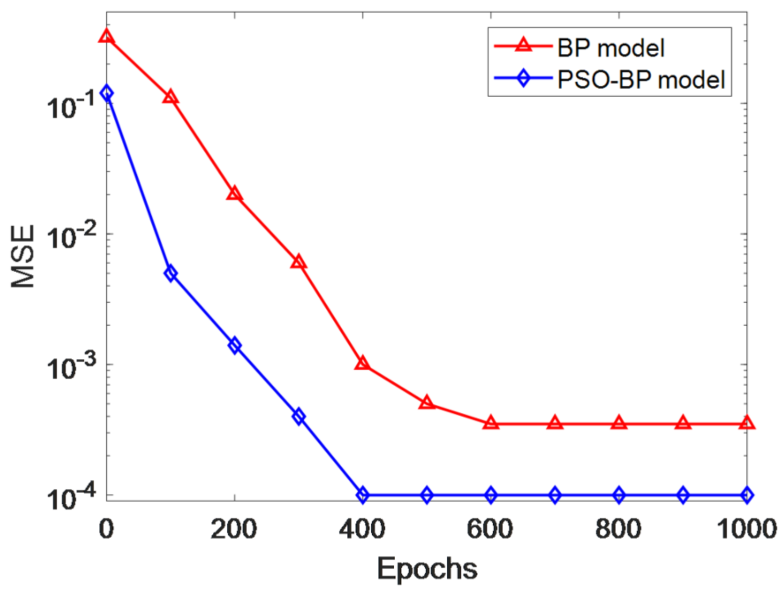


Table 2. Optimization effect of different particle numbers on neural network model

Number of Particles	Number of Iterations	Training Sample Average Error
5	245	0.05
10	435	0.25
20	290	0.16

Figure 4. Iteration times curves of different models



Implementation Path of Corpus English Translation Based on Blockchain

1) Create a "blockchain + AI" English translation corpus. Blockchain can integrate foreign language and corpus-related resources to create an English translation sharing platform. Trade across borders is becoming more common as the "One Belt One Road" plan advances, necessitating a greater need for bilingual professionals fluent in English and their native language. In order to address the demands of compound talent, a platform built on the blockchain might bring together academics, business leaders, and members of the general public. Blockchain includes three forms: public chain, private chain, and alliance chain. Based on blockchain and AI technology, we will encourage financial enterprises, banking associations, financial institutions, and other institutions to practice English translation teaching to build a public chain-based decentralized network for sharing educational resources. Students on the alliance chain platform may now benefit from personalized instruction and hands-on experience in English translation as the platform's primary student English translation professional module.

Corpus refers to a collection of texts, including written and spoken language, of a certain scale of authentic language material stored and processed by a computer. The real corpus collected by the corpus has become a vital language teaching and research resource. Using blockchain and AI technology, based on building an English translation-sharing platform, an English translation corpus is built to provide intelligent support for English translation practice, teaching, and research (Xiang & Liu, 2018). All English professional colleges unite to create a shared and co-constructed English translation corpus. Each university can upload the corpus collected in the translation teaching practice of the university in time and jointly maintain the English translation corpus (Tran, 2022). In practical teaching, teachers and students carry out teaching activities according to the corpus's vocabulary, terminology, word choice and sentence construction, text structure, and translation samples. In translation studies, linguistics and translation researchers use corpus resources to conduct natural and social sciences, such as neural processing and computer processing of language forms and information, machine translation model construction, and the impact of cultural practices on translation.

2) Improve the machine English translation software of the "blockchain + AI" English translation training platform. With the rapid development of AI technology, the AI language service market is active, and the AI translation technology is constantly updated and iterated. People favor AI translation because of its low cost and high efficiency. AI supports text translation and performs additional tasks, such as character recognition, image recognition, and speech synthesis, to achieve translation tasks in some situations. For example, Sogou Translation, with its powerful search engine and massive resources, can realize text, photo, and voice translation functions and machine simultaneous interpreting translation activities. However, AI translation cannot wholly replace human translation because improving AI translation technology depends on the number of existing corpora and solid professional basic knowledge. The large community, English translation-sharing platform, and corpuses built by blockchain technology lay the foundation for improving the quality of machine translation software. Translation that machines cannot complete due to context and cultural differences or mistranslation caused by translation deviation can be completed by manual proofreading. Big data recommended by the community can form the foundation for the blockchain based on the qualifications, resumes, and preferences of the community personnel, and the compilation and unification work is completed collaboratively. Aside from saving money on education expenditures, using English translation software to conduct man-machine simulation exercises increases students' translation capacity.

The English translation training exercise, the English translation sharing platform, the English translation sharing corpus, and the machine English translation software jointly create an English translation training exercise platform. The AI technology of translation and the blockchain big data technology can improve the translator's ability to integrate English resources and facilitate the construction of the English translation curriculum system. As a result, the required, optional, and inclusive courses in the English translation program are thoughtfully organized, dispersed, and visually represented, following the course goals, content, and instructors. English translation abilities are needed to accomplish the goals of harmonizing educational material and teaching standards. Relevant enterprises, translation companies, and universities use big data and AI technology when conducting English translation training and practical teaching activities to automatically screen relevant data from the platform according to individual needs, preferences, and expertise and provide real-time recommendations to students. Enterprises and universities can also use AI technology for other tasks, such as extracting data, designing context, creating on-site simulated translation, and conducting man-machine dialogue, to provide intelligent support for English translation training of enterprises and schools and meet the self-training of translators inside and outside the classroom.

Combining the mining function of blockchain and the automatic calculation of artificial intelligence can improve the translation ecological chain. The consensus mechanism of blockchain is to get paid according to work and get more for more work. The computer intelligently calculates the translator's computational power and time length according to the value of the contribution and the task completed to determine the translator's workload, thus obtaining a particular reward, namely, a virtual token. For business translation and privacy, features include business contracts, securities transactions, and deposit certificates. Use the encryption technology of blockchain to realize the security of information transactions. Before sending the information to the designated translator, the user uses the public key to encrypt the information. Only the translator with the private key can decrypt the information. Commercial translation can achieve confidential translation, avoid transaction disclosure, and achieve information transaction security.

CONCLUSION

The effective combination of blockchain and AI will help improve the efficiency and quality of English translation and promote the deep integration of English translation practice and teaching. This paper first introduces the research status of foreign language corpus and blockchain English translation in China. Then, it introduces the basic principles of BPNN and particle swarm optimization and constructs the PSO-BP model.

Experimental results show that the prediction accuracy of BPNN optimized by particle swarm optimization algorithm is greatly improved, the convergence speed is faster, and it will not fall into the local optimal trap. Finally, this paper proposes the implementation path of blockchain in corpus translation application: (1) build "blockchain+ AI" English translation corpus; (2) Improve the machine English translation software of the "blockchain+ AI" English translation training platform. While consolidating the theoretical foundation of translation, translators also need to constantly learn the cutting-edge knowledge of blockchain and AI technology in translation corpus and teaching and jointly promote the development of translation and teaching towards specialization and intellectualization from aspects such as talent training, technology research and development, and industrial development.

REFERENCES

Baker, M. (2004). A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics*, *9*(2), 167–193. doi:10.1075/ijcl.9.2.02bak

Baker, M. (2019). Corpus-based translation studies: The challenges that lie ahead. In K. H. Kim & Y. Zhu (Eds.), *Researching translation in the age of technology and global conflict* (pp. 44–54). Routledge. doi:10.4324/9780429024221-4

Baker, M. (2019). Corpus linguistics and translation studies: Implications and applications. In K. H. Kim & Y. Zhu (Eds.), *Researching translation in the age of technology and global conflict* (pp. 9–24). Routledge. doi:10.4324/9780429024221-2

Bowker, L. (2001). Towards a methodology for a corpus-based approach to translation evaluation. *Meta*, *46*(2), 345–364. doi:10.7202/002135ar

Brynjolfsson, E., Hui, X., & Liu, M. (2019). Does machine translation affect international trade? Evidence from a large digital platform. *Management Science*, *65*(12), 5449–5460. doi:10.1287/mnsc.2019.3388

Busari, S. A., Huq, K. M. S., Mumtaz, S., Rodriguez, J., Fang, Y., Sicker, D. C., Al-Rubaye, S., & Tsourdos, A. (2019). Generalized hybrid beamforming for vehicular connectivity using THz massive MIMO. *IEEE Transactions on Vehicular Technology*, *68*(9), 8372–8383. doi:10.1109/TVT.2019.2921563

Du, J., Jiang, C., Han, Z., Zhang, H., Mumtaz, S., & Ren, Y. (2017). Contract mechanism and performance analysis for data transaction in mobile social networks. *IEEE Transactions on Network Science and Engineering*, *6*(2), 103–115. doi:10.1109/TNSE.2017.2787746

Duan, J., & Wei, J. (2021). A genre analysis of English and Chinese legal research article abstracts: A corpus-based approach. *Journal of Language Teaching and Research*, *12*(5), 810–821. doi:10.17507/jltr.1205.21

Duro Moreno, M. (2020). Translation quality gained through the implementation of the iso en 17100: 2015 and the usage of the blockchain: The case of sworn translation in Spain. *Babel*, *66*(2), 226–253. doi:10.1075/babel.00153.mor

Green, S., Heer, J., & Manning, C. D. (2015). Natural language translation at the intersection of AI and HCI. *Communications of the ACM*, *58*(9), 46–53. doi:10.1145/2767151

Inwood, O., & Zappavigna, M. (2021). Ideology, attitudinal positioning, and the blockchain: A social semiotic approach to understanding the values construed in the whitepapers of blockchain start-ups. *Social Semiotics*, *33*(3), 451–469. doi:10.1080/10350330.2021.1877995

Jiang, F. (2020). The reform of cultivation mode of Chinese university english translation talents in the age of artificial intelligence. *Higher Education of Social Science*, *18*(1), 45–49. doi:10.3968/11613

Jung, Y., & Agulto, R. (2019). Integrated management of network address translation, mobility and security on the blockchain control plane. *Sensors (Basel)*, *20*(1), 69. doi:10.3390/s20010069 PMID:31877695

Lin, S., Zheng, B., Alexandropoulos, G. C., Wen, M., Chen, F., & Mumtaz, S. (2020). Adaptive transmission for reconfigurable intelligent surface-assisted OFDM wireless communications. *IEEE Journal on Selected Areas in Communications*, *38*(11), 2653–2665. doi:10.1109/JSAC.2020.3007038

Lyons, G., & Urry, J. (2005). Travel time use in the information age. *Transportation Research Part A, Policy and Practice*, *39*(2-3), 257–276. doi:10.1016/j.tra.2004.09.004

Pan, B., & Qin, Q. (2022). Construction of parallel corpus for english translation teaching based on computer aided translation software. *Computer-Aided Design and Applications*, *19*(1), 70–80. doi:10.14733/cadaps.2022.S1.70-80

Peterlin, A. P. (2010). Hedging devices in Slovene-English translation: A corpus-based study. *Nordic Journal of English Studies*, *9*(2), 171–193. doi:10.35360/njes.222

Shlesinger, M. (1998). Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta*, *43*(4), 486–493. doi:10.7202/004136ar

Sukman, K., Triwatwaranon, W., Munkongdee, T., & Chumnumnawin, N. (2022). A corpus-based study of lexical collocations of keywords found in online business news articles. *European Journal of English Language Teaching*, *7*(3), 59–68. doi:10.46827/ejel.v7i3.4275

Tran, T. (2022). Predicting digital asset prices using natural language processing. *Survey (London, England)*. Advance online publication. doi:10.48550/arXiv.2212.00726

Wan, Z., Xia, X., & Hassan, A. E. (2021). What is discussed about blockchain? A case study on the use of balanced lda and the reference architecture of a domain to capture online discussions about blockchain platforms across the stack exchange communities. *IEEE Transactions on Software Engineering*, *47*(7), 1331–1349. doi:10.1109/TSE.2019.2921343

Wang, X. (2021). Building a parallel corpus for English translation teaching based on computer-aided translation software. *Computer-Aided Design and Applications*, *18*(S4), 175–185. doi:10.14733/cadaps.2021.S4.175-185

Wei, Y., Wen, J., Huang, J., & Jiang, S. (2022). Research on translation skills of online literature and cultivation of translation talents. *International Journal of Social Science and Education Research*, *5*(2), 242–248. doi:10.6918/IJOSSER.202202_5(2).0039

Xiang, D., & Liu, C. (2018). The semantics of MOOD and the syntax of the Let's-construction in English: A corpus-based Cardiff Grammar approach. *Australian Journal of Linguistics*, *38*(4), 549–585. doi:10.1080/07268602.2018.1510726

Zanettin, F. (2013). Corpus methods for descriptive translation studies. *Procedia: Social and Behavioral Sciences*, *95*, 20–32. doi:10.1016/j.sbspro.2013.10.618

Zhang Aiping, male, Shaoyang, Hunan, born in November 1983, member of the Communist Party of China, Master's degree, a lecturer in the School of Foreign Language from Changsha Social Work College.He has been engaged in teaching in colleges for many years, published academic papers in publications above the provincial level, and participated in a number of provincial-level projects. His main research direction includes Blended Teaching of Higher Vocational English, English Translation, and Huxiang Culture.

Zhuxiaowei Zhu, male, born in Shiyan, Hubei Province in August, 1982, is a member of the Communist Party of China, a master's student, an associate professor of the school of foreign languages and foreign trade of Wuhan Polytechnic, and his research direction is higher vocational education.