Design of English Grammar Error Correction System Based on Deep Learning Algorithms

Wei Cao*

{*Corresponding author: caowei66@21cn.com}

Sichuan University of Media and Communications, Chengdu, Sichuan Province, China

Abstract. With the popularity of the Internet, English has become a globally used language, and English grammar(EG) errors have become a universal problem worldwide. EG errors not only affect daily communication but also cause reading and writing mistakes. Traditional grammar analysis methods mainly look for EG errors through static analysis, but this method can only detect some fixed types of grammar errors and cannot provide real-time EG correction. Based on this, this paper proposes an EG correction system based on deep learning(DL) algorithms. Through this system, it is hoped to provide more accurate and real-time EG error detection services. According to the experimental results of the system, it can be concluded that the grammar error recognition accuracy of the sequence model is the highest, reaching 82.2%, but its recall rate is slightly lower, only 42.3%. The accuracy of the grammar correction system based on DL can be further improved. The design, implementation, and testing processes of the system provide valuable references for the practical application of DL technologies in natural language processing tasks, such as grammar error detection and correction.

Keywords: Deep Learning, English Grammar, Error Correction System, Sequence Model

1 Introduction

With the development of globalization and the spread of information technology, English has become a universally used language. However, many people from non-English speaking countries often make grammar mistakes when learning English, which not only affects communication but also impacts their enthusiasm for learning. Therefore, it is extremely significant to develop an efficient EG correction system. DL algorithms have been widely used as powerful tools in the field of natural language processing in recent years. This paper will explore the design of an EG correction system based on DL algorithms, aimed at improving the grammar level of English learners and the effectiveness of communication.

In traditional grammar correction systems, an important feature is structure reconstruction. DL algorithms can effectively solve this problem. Grammar error correction(EC) technology based on DL is a new research direction. Scholars have conducted detailed analysis and evaluation of the EG correction system based on DL

ICSETPSD 2023, November 17-18, Coimbatore, India

DOI 10.4108/eai.17-11-2023.2342643

algorithms. DL algorithms have strong learning and generalization capabilities and can effectively solve many problems in natural language processing [1-2]. In the study of the design and implementation of EG correction system based on DL, some researchers have proposed a design scheme for an EG correction system based on DL algorithms. This system can meet the needs of EG correction [3-4]. Others have summarized and evaluated the application of DL algorithms in the field of natural language processing. They believe that DL algorithms have become the mainstream algorithms in the field of natural language processing, and they have a wide range of applications in text classification, sentiment analysis, machine translation, question-answering systems, etc. [5-6]. This paper will explore the design of an EG correction system based on DL algorithms, mainly to improve the grammar level and communication effectiveness of English learners.

This paper mainly researches the EG correction system based on DL algorithms. It first elucidates the definition and classification of EG errors. By classifying EG errors, we can better understand which grammar errors need to be paid close attention to. It then explores the application of DL algorithms in EG correction. By combining DL algorithms with other technologies, an EG correction system is constructed. Finally, it studies the EG correction system based on machine learning. In this paper, we will explore how to apply machine learning technology to the EG correction system to achieve more accurate and real-time EG error detection.

2 English Grammar Error Correction Method Based on Deep Learning

2.1 English grammar and Syntax Error

EG is a very important part of English learning, as it is the foundation and core of the English language. Grammar rules determine the structure and meaning of English sentences, and mastering EG can help students understand and express English correctly, avoiding language misunderstandings and ambiguity [7]. In English exams, grammar is also a very important assessment content, and mastering EG can improve exam scores [8].

Grammar errors refer to mistakes made in the use of EG. Part-of-speech errors usually involve the incorrect use of nouns, verbs, adjectives, etc. Tense errors refer to the misuse of verb tenses, often manifested in inconsistencies and confusions in tense. Voice errors refer to the misuse of passive and active voice. Sentence structure errors, such as subject-verb disagreement or mismatch between the subject and the predicate, mainly point to errors in sentence structure. Punctuation errors, as the name suggests, are mistakes in the use of punctuation, typically due to improper use.

It can use several methods in Figure 1 to find English Syntax error:

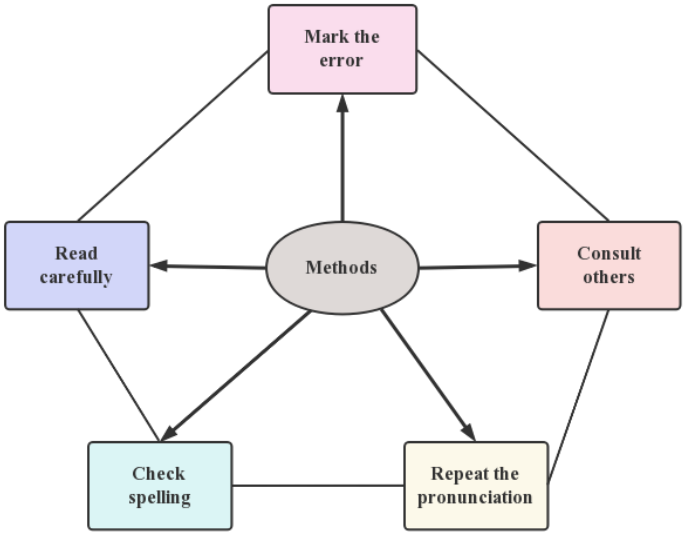

Fig.1 Ways to find EG errors

When reading English articles, one needs to carefully read each word and sentence, especially paying attention to some common grammar mistakes. You can mark potential grammar errors with a pen for further checking and correction later. When checking for grammar errors, you also need to ensure the spelling is correct, as spelling mistakes can also affect the use of grammar. When checking for grammar errors, you can try to read the article out loud, making it easier to spot errors like repeated sounds or improper tones. If you encounter errors that you can't identify, you can seek help from teachers or other English proficient individuals and listen to their advice and feedback. Learning EG is very important for English learning, and grammar errors are common mistakes in English usage that require special attention.

2.2 Deep Learning Algorithm and Its Application in Grammar Error Correction

Grammar correction plays a crucial role in English learning. The primary method to improve English teaching quality and reduce confusion and misunderstanding from the mother tongue is to correct grammar errors based on language rules and patterns. However, these methods usually require fine-grained programming and transferring of language knowledge bases, and it's challenging to cover all possible errors. DL, as a machine learning technology, can effectively solve this problem[9]. Computers can automatically learn from big data and discover hidden information relationships between data. Some grammar EC models based on DL are becoming increasingly popular and are being used more and more in the field of natural language processing[10].

DL is widely used in areas such as image recognition, speech analysis, and natural language processing[11-12]. It employs multi-level neural networks to learn functions in data, constructing models that can independently extract common expressions from

input data, and continually adjusts through feedback to improve model accuracy[13]. Grammar correction refers to the elimination of certain spelling, grammar, or other writing errors to make the text more conforming to the context language rules. The most common grammar errors include inconsistent main headings, misuse of adjectives and adverbs, etc. Currently, many grammar EC models based on DL can quickly identify and correct simple mistakes.

Currently, one of the most important methods of grammar correction is based on language models. This method analyzes the mathematical probabilities of input sets and then compares them with other candidate sets to obtain a more accurate set. Google's grammar assistant uses this method, employing a statistical model known as an n-gram language model. Although this method based on language models has achieved good results in solving grammar errors, it has some limitations. When dealing with specific languages and terms beyond the scope of training data, this method may become less effective. Another grammar correction method is rule-based, which uses predefined grammar rules and regular expressions to match them. The disadvantage of this method is that it requires manually writing many rules and highly depends on different languages. Although the accuracy of rule-based methods is acceptable, the requirements for input data are high, and developing rules takes a long time. The success of DL in speech recognition, image recognition, and natural language processing shows its great potential [14-15].

2.3 Deep Learning Grammar Error Correction Model

With the development of artificial intelligence technology, DL has become an important technique in the field of natural language processing[16]. DL models for EG correction can assist students and non-native English speakers to better grasp EG knowledge and improve the accuracy and fluency of EG usage.

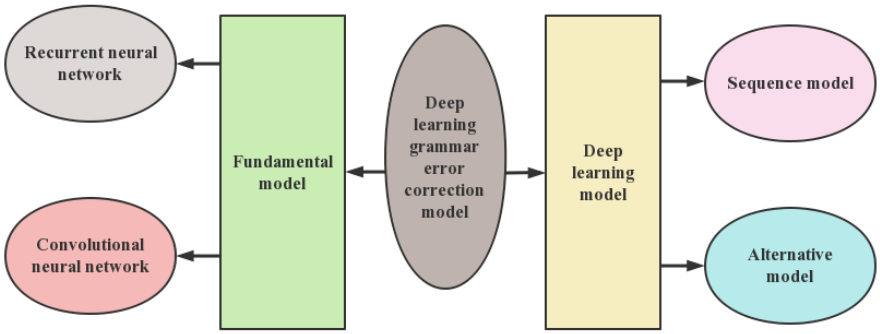

Fig.2 DL grammar error correction model

Before creating a DL model for EG correction, data preprocessing is required. The corpus forms the basis of the DL model, so collecting a large number of EG errors and correct examples is crucial. There may be many errors, unnecessary or non-standard data in the collected corpus, hence data cleaning is needed. Data cleaning includes operations such as deleting tags, removing special symbols, and deleting unnecessary information. The composition of the DL grammar correction model is shown in

Figure 2:

Recursive Neural Networks and Convolutional Neural Networks are two common model structures in DL[17-18]. Recursive Neural Networks are a type of sequence model that can process sequence data of any length while retaining previous state information. In grammar correction, it can predict and correct the current text by learning the content of the previous text. Convolutional Neural Networks, a model suitable for image processing, can also be applied in the field of natural language processing. In grammar EC, convolution operations can be used to capture local features in the text and pooling operations can be used to reduce the number of model parameters. In addition to the basic model structure, there are improved model structures that can be applied to EG EC models. Sequence models are a type of DL model that can process sequence data. In grammar correction, sequence models can improve correction capability through learning to better handle text sequences. The generation algorithm used by sequence models is as follows:

$$q(B|A|) = \sum_{i=1}^{M} q(b_i|B_{\prec i}, A; \Theta_{T2T}) \quad (1)$$

The final Loss function is:

$$L_{T2T}(B|A; \Theta_{T2T}) = -\sum_{i=1}^{M} \log q(b_i|B_{\prec i}, A; \Theta_{T2T}) \quad (2)$$

Among them, $B_{\prec i}$ refers to the first i-1 character of the target sentence during the training phase. The models used to generate editing operations are mostly sequence annotation class models. And the entire model can be represented as:

$$q(F|A) = q((f_1, f_2, \cdots, f_M)|A; \Theta) \quad (3)$$

The transformation model is an emerging DL model. In grammar correction, transformation models can learn the relationships between text sequences through self-attention mechanisms, thereby better handling deep dependencies in text. The English Syntax EC model of DL can help people better master EG skills and improve the accuracy and fluency of grammar [19-20]. By continuously optimizing the structure and parameters of the model, the EC effect of the model can be continuously improved.

2.4 Corpus Processing and Error Correction Algorithms

Corpus processing is an essential aspect of natural language processing, which primarily includes sentence segmentation, tokenization, language tagging, and corpus techniques. Sentence segmentation is the process of splitting a piece of text into multiple sentences. In English, punctuation is more flexible and must be combined with grammar rules and context to form clauses. Sentence segmentation can help better understand the semantics of a text, which is very important for natural language processing. Tokenization is the process of breaking text into several words. Given the

absence of clear delimiters in English characters, it is necessary to use tokenization techniques to divide the text into words. Tokenization is essential for tasks such as machine translation and information retrieval. Language tagging, usually involving nouns, verbs, adjectives, adverbs, prepositions, etc., is the process of assigning the corresponding part of the language title to each word. Some language tagging techniques are of great significance to natural language processing. A corpus refers to a collection of a large amount of organized text data, typically used for application research in natural language processing. A balanced corpus is a collection of text data from different sources, fields, and time periods to ensure the diversity and balance of the text data. An unbalanced corpus refers to text data collected from a specific area or time period.

Grammar EC refers to the process of detecting and correcting grammar errors in text using natural language processing techniques. Grammar-based EC includes using syntactic analysis techniques to analyze text, detect grammar errors, and provide appropriate correction suggestions. Syntactic analysis is performed to obtain the grammatical structure of the text, then the grammatical structure is checked to identify errors and make appropriate corrections. Grammar-based correction methods can deeply analyze and handle grammatical errors, but due to the complexity of grammatical analysis techniques, they work slowly. Rule-based grammar correction involves using manually defined rules to analyze text, detect grammar errors, and provide appropriate correction suggestions. This process first defines some rules for detecting grammatical errors in the text, then analyzes the text to identify errors and suggest appropriate corrections. Rule-based grammar EC methods have the advantage of fast execution and easy implementation, but due to rule restrictions, their error detection and correction capabilities are limited to a certain extent.

Corpus processing and grammar correction are crucial techniques in the field of natural language processing and hold significant importance in enhancing the efficiency and accuracy of natural language processing systems. It is necessary to continuously optimize and improve these techniques to achieve better natural language processing results.

2.5 The English Grammar Error Correction System

The grammar correction system is modeled and divided into modules according to business function requirements. To achieve loose coupling, easy extension, and reusability between modules, the system is divided into modules according to business boundaries. The frontend module is responsible for rendering the page and displaying logic. The grammar correction module is the core module of the system, mainly consisting of three functions: data processing, model training, and model correction. The service access module is the system's backend management system. The feedback filtering module implements feedback filtering algorithms and provides feedback filtering services. The main functions of the backend module include model training, grammar correction, feedback suggestions, user login, and other functions. The grammar correction module provides two main external interfaces: grammar correction and model training.

The grammar correction interface is a function to correct the grammar of the input English sentence. Its main workflow involves sentence segmentation, using the well-trained correction model for correction, and then summarizing the correction

results for return. The architecture of the grammar correction system is shown in Figure 3.

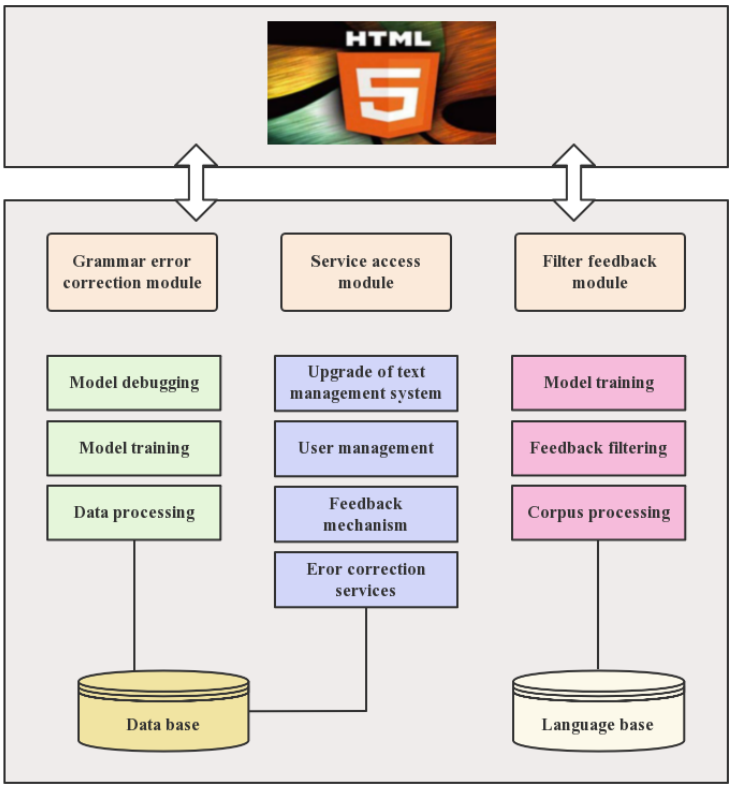

Fig.3 Syntax-EC system architecture

The grammar correction module provides a model training interface, which can be used by external systems to trigger re-training of the model to improve the effect of grammar correction, which is an essential part of system self-updating. Before training starts, the corpus is regularized using corpus preprocessing scripts, and then model initialization training is conducted. If any exceptions occur during corpus preprocessing, model initialization, or model training, a training exception notification is sent to the administrator for system repair.

3 Experimental Design of English Grammar Error Correction System

3.1 Experimental Design

First, it would need to collect a large number of Syntax error data, including sentences, paragraphs, articles, etc. These data can be obtained through web crawler, manual

annotation or other methods. For the collected data, preprocessing such as removing noise and filtering useless information is necessary, such as removing noise and filtering useless information. At the same time, it is also necessary to convert the data into a format suitable for DL models. Choose an appropriate DL model. Different model structures can be selected based on the scale and characteristics of the data. Use the preprocessed data to train the model, and continuously optimize the model parameters through the Backpropagation, so that the model can better correct Syntax error. Evaluate the trained model using a test set, calculate the accuracy, recall, and other indicators of the model, as well as error analysis. Integrate the trained model into a grammar correction system to achieve automated grammar correction functions. Optimize the grammar correction system for practical application scenarios, incorporate user feedback mechanisms, and add grammar rules.

3.2 Building a System Development Environment and System Operation

This study requires the use of a high-performance computer or server to ensure the training of the DL model and the speed of inference. Install the DL framework, Python programming environment, and other necessary dependency libraries to support model training and system development. According to the designed process, the various modules of the grammar correction system are developed in sequence, including data collection, data preprocessing, model training, model evaluation, system integration, etc. The developed grammar correction system needs to be tested, including functionality testing, performance testing, stability testing, etc., to ensure that the system can run normally and meet requirements. The developed grammar correction system is deployed in actual application scenarios.

3.3 Optimization of English Grammar Error Correction System

Adding a user feedback mechanism to a grammar correction system can help the system automatically learn more grammar rules and error types, improving the effectiveness of EC. Adding more grammar rules to the system can cover more types of Syntax error and improve the accuracy of EC. According to the actual application scenarios and data characteristics, optimize the DL model, such as adding hidden layers, adjusting the Learning rate, etc., to improve the generalization ability and performance of the model. By using data augmentation technology, the diversity of training data is increased, and the robustness and generalization ability of the model are improved. For multilingual environments, it is necessary to optimize grammar rules for different languages to improve grammar EC effectiveness. At the same time, it is also necessary to collect and preprocess multilingual data.

3.4 Test Results

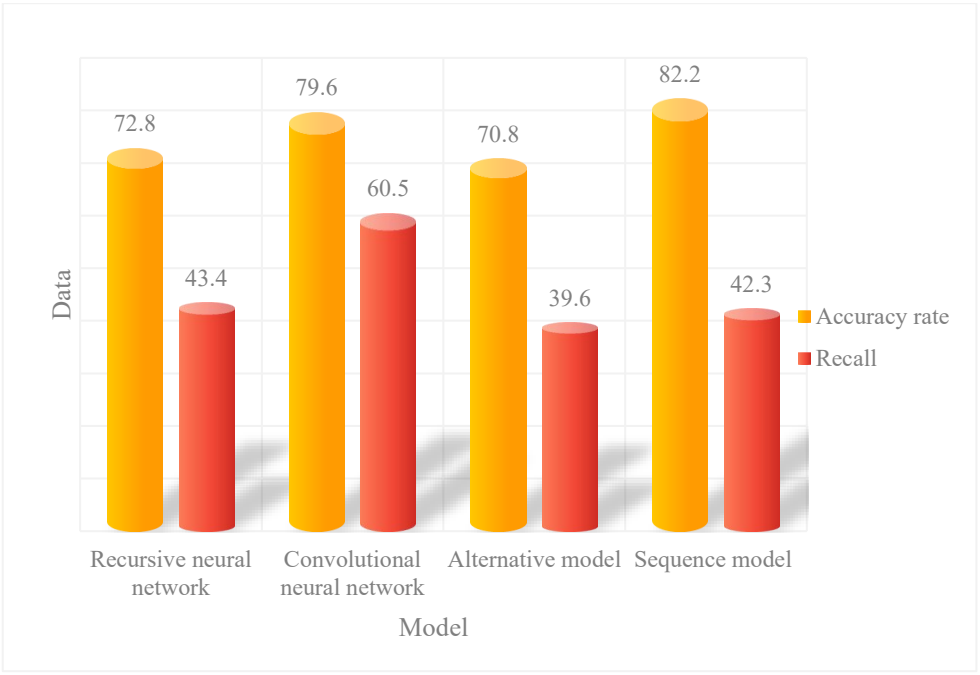

Fig.4 Accuracy and recall for the different models

As shown in Figure 4, the accuracy of recurrent neural networks is relatively low, but compared to the transformation model, its accuracy is 2 percentage points higher. The accuracy of Convolutional neural network ranked second, but the recall rate ranked first. The accuracy of the conversion model reached 70.8%, with the lowest recall rate of 39.6%.

As shown in Table 1, the functional test results of the EG correction system mentioned in this article show that the error rate of the correction function is 2%, with a concurrent number of 382. The error rate of page display is 0.4%, with a concurrent number of 450. The error rate of feedback suggestions is 6.5%, with a concurrent number of 320. The error rate of user login is 1%, with a concurrent number of 500.

Table 1. Functional test results of the EG error correction system

	Error rate	Concurrent number
EC function	2	382
Page display	0.4	450
Feedback advice	6.5	320
Users login	1	500

This article conducted experiments on adapters of different sizes. In these experiments, most of the hyperparameters of the two-stage EC system were fixed, and the accuracy of the EC system using adapters of different sizes is shown in Figure 5:

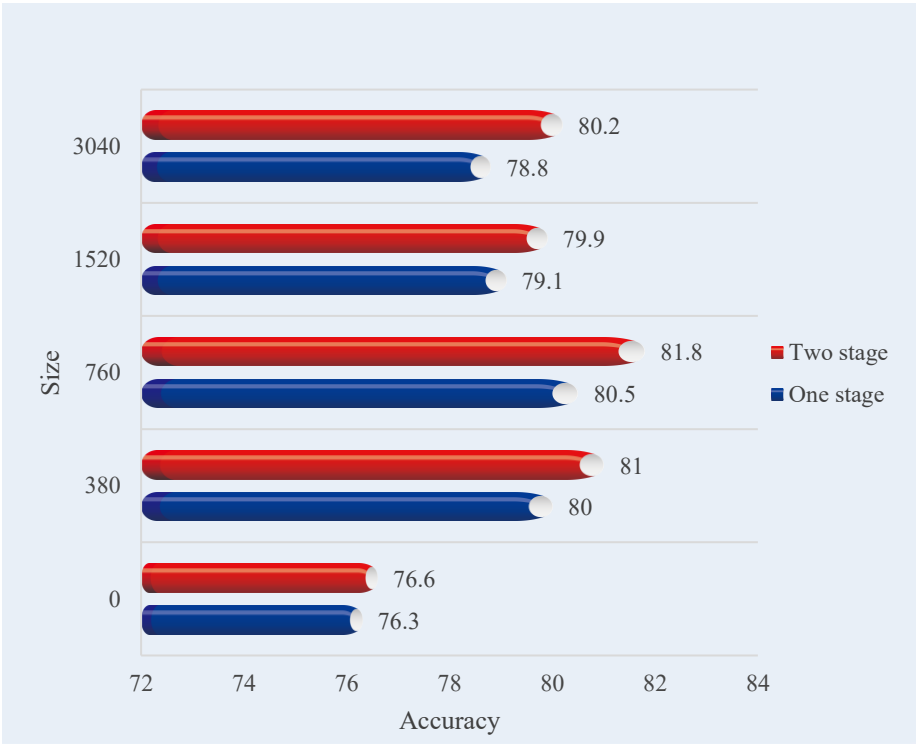

Fig.5 Accuracy of adapter EC systems of different sizes

As shown in Figure 5, when the internal feature dimension of the adapter is set to 760, the EC system achieves the best EC results. When the internal feature dimension is less than 760, the larger the feature dimension, the better the EC system results. On the contrary, when it is greater than 760, its EC effect would decrease. In addition, it can be observed from the graph that the accuracy of the two stages is higher and the effect is better.

4 Application and Prospect of English Grammar Error Correction System

EG correction based on DL algorithms has a wide range of application scenarios in practical use and a very broad development prospect. Online writing tools are typical application scenarios for DL grammar correction. By annotating and correcting grammar errors in real-time during the writing process, the system can help users improve writing efficiency and the quality of articles. Academic research is a type of text that strictly follows grammar rules; the system can assist scholars in checking for grammar errors in academic research, further improving the quality of papers and the pass rate of article reviews. The grammar correction feature in mobile applications can help users better express themselves in scenarios such as social media and text

messages, avoiding embarrassment caused by grammatical errors. In English education and training, the system can help students discover and correct grammar errors in a timely manner, thereby improving English learning outcomes.

With the acceleration of globalization, the demand for grammar correction in a multilingual environment is increasing. DL grammar correction technology will support EC in multiple languages to meet the needs of different countries and regions. The system will have stronger self-learning ability and can improve correction results and accuracy through self-learning and optimization. At the same time, the system will combine natural language processing technology to achieve more intelligent grammar correction functions, thus improving correction accuracy. The system will also support personalized customization; based on the user's language habits and types of grammar errors, customized correction plans can be made to improve correction effects and user satisfaction.

EG correction technology based on DL algorithms will further develop and be applied, providing more accurate and fluent support for people's language expression.

5 Conclusions

This paper proposes a design scheme for an EG correction system, aiming to improve the grammar level and communication effectiveness of English learners. The system was evaluated through experiments, and it is capable of providing more accurate and real-time EG error detection services. The grammar correction system proposed in this paper has a certain degree of innovation and practicality, and it can effectively improve the accuracy and real-time performance of EG error detection. However, follow-up research needs to further address the system's stability and data shortage issues in order to achieve better results.

References

[1] Alvaro Fernandez-Quilez: Deep learning in radiology: ethics of data and on the value of algorithm transparency, interpretability and explainability. AI Ethics 3(1): 257-265 (2023)

[2] Nesrine Kadri, Ameni Ellouze, Mohamed Ksantini, Sameh Hbaieb Turki: New LSTM Deep Learning Algorithm for Driving Behavior Classification. Cybern. Syst. 54(4): 387-405 (2023)

[3] Leah Gosselin, Clara D. Martin, Ana González Martín, Sendy Caffarra: When A Nonnative Accent Lets You Spot All the Errors: Examining the Syntactic Interlanguage Benefit. J. Cogn. Neurosci. 34(9): 1650-1669 (2022)

[4] Nankai Lin, Boyu Chen, Xiaotian Lin, Kanoksak Wattanachote, Shengyi Jiang: A Framework for Indonesian Grammar Error Correction. ACM Trans. Asian Low Resour. Lang. Inf. Process. 20(4): 57:1-57:12 (2021)

[5] Samuel Sousa, Roman Kern: How to keep text private? A systematic review of deep learning methods for privacy-preserving natural language processing. Artif. Intell. Rev. 56(2): 1427-1492 (2023)

[6] Abdul Rahaman Wahab Sait, Mohamad Khairi Ishak: Deep Learning with Natural Language Processing Enabled Sentimental Analysis on Sarcasm Classification. Comput. Syst. Sci. Eng. 44(3): 2553-2567 (2023)

[7] Alexander Zabolotskikh, Anna Zabolotskikh, Tatiana Dugina, Daria Tavberidze: Creating individual learning paths in the Moodle plugin for undergraduate students to study English grammar. Educ. Inf. Technol. 26(1): 617-637 (2021)
[8] Ivana Simonova: Blended approach to learning and practising English grammar with technical and foreign language university students: comparative study. J. Comput. High. Educ. 31(2): 249-272 (2019)
[9] Erdal Özbay: An active deep learning method for diabetic retinopathy detection in segmented fundus images using artificial bee colony algorithm. Artif. Intell. Rev. 56(4): 3291-3318 (2023)
[10] S. Prasanna Bharathi, S. Srinivasan, G. Chamundeeswari, B. Ramesh: Hybrid Deep Learning-Improved BAT Optimization Algorithm for Soil Classification Using Hyperspectral Features. Comput. Syst. Sci. Eng. 45(1): 579-594 (2023)
[11] M. Muthulakshmi, G. Murugeswari: A hybrid grey wolf-whale optimization algorithm for classification of corona virus genome sequences using deep learning. Int. Arab J. Inf. Technol. 20(3): 331-339 (2023)
[12] Sheela J, Janet B: Caviar-Sunflower Optimization Algorithm-Based Deep Learning Classifier for Multi-Document Summarization. Comput. J. 66(3): 727-742 (2023)
[13] V. T. Ram Pavan Ku, M. Arulselvi, K. B. S. Sastry: An Optimized Deep Learning Based Optimization Algorithm for the Detection of Colon Cancer Using Deep Recurrent Neural Networks. Int. J. Commun. Networks Inf. Secur. 14(1s): 22-36 (2023)
[14] S. Sathyavathi, K. R. Baskaran: An Intelligent Human Age Prediction from Face Image Framework Based on Deep Learning Algorithms. Inf. Technol. Control. 52(1): 245-257 (2023)
[15] M. J. Jeyasheela Rakkini, K. Geetha: Deep learning classification of bitcoin miners and exploration of upper confidence bound algorithm with less regret for the selection of honest mining. J. Ambient Intell. Humaniz. Comput. 14(6): 6545-6561 (2023)
[16] Anna M. Díaz-Rovira, Helena Martín, Thijs Beuming, Lucía Díaz, Victor Guallar, Soumya S. Ray: Are Deep Learning Structural Models Sufficiently Accurate for Virtual Screening? Application of Docking Algorithms to AlphaFold2 Predicted Structures. J. Chem. Inf. Model. 63(6): 1668-1674 (2023)
[17] Moruf Akin Adebowale, Khin T. Lwin, M. Alamgir Hossain: Intelligent phishing detection scheme using deep learning algorithms. J. Enterp. Inf. Manag. 36(3): 747-766 (2023)
[18] Ilyas Ozer, Caner Kaya Ozer, Ali Can Karaca, Kutlucan Gorur, Ismail Kocak, Onursal Cetin: Species-level microfossil identification for globotruncana genus using hybrid deep learning algorithms from the scratch via a low-cost light microscope imaging. Multim. Tools Appl. 82(9): 13689-13718 (2023)
[19] Debabrata Dansana, Raghvendra Kumar, Aishik Bhattacharjee, D. Jude Hemanth, Deepak Gupta, Ashish Khanna, Oscar Castillo: Early diagnosis of COVID-19-affected patients based on X-ray and computed tomography images using deep learning algorithm. Soft Comput. 27(5): 2635-2643 (2023)
[20] Mohammad Bany Taha, Chamseddine Talhi, Hakima Ould-Slimane, Saed Alrabaee, Kim-Kwang Raymond Choo: A Multi-Objective Approach Based on Differential Evolution and Deep Learning Algorithms for VANETs. IEEE Trans. Veh. Technol. 72(3): 3035-3050 (2023)