Research Article

English Grammar Error Correction Algorithm Based on Classification Model

Shanchun Zhou¹ and Wei Liu²

¹*School of Foreign Languages, Xinyang College, Xinyang, Henan 464000, China*

²*Intercultural Language Research Center, Xinyang Normal University, Xinyang, Henan 464000, China*

Correspondence should be addressed to Wei Liu; xylw@xynu.edu.cn

Received 13 December 2020; Revised 6 January 2021; Accepted 8 January 2021; Published 19 January 2021

Academic Editor: Wei Wang

Copyright © 2021 Shanchun Zhou and Wei Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

English grammar error correction algorithm refers to the use of computer programming technology to automatically recognize and correct the grammar errors contained in English text written by nonnative language learners. Classification model is the core of machine learning and data mining, which can be applied to extracting information from English text data and constructing a reliable grammar correction method. On the basis of summarizing and analyzing previous research works, this paper expounded the research status and significance of English grammar error correction algorithm, elaborated the development background, current status, and future challenges of the classification model, introduced the methods and principles of feature extraction method and dynamic residual structure, constructed a basic model for English grammar error correction based on the classification model, analyzed the classification model and translation model of English grammar error correction, proposed the English grammar error correction algorithm based on the classification model, performed the analyses of the model architecture and model optimizer of the grammar error correction algorithm, and finally conducted a simulation experiment and its result analysis. The study results show that, with the continuous increase of training samples and the continuous progress of learning process, the proposed English grammar error correction algorithm based on the classification model will continue to increase its classification accuracy, further refine its recognition rules, and gradually improve correction efficiency, thereby reducing processing time, saving storage space, and streamlining processing flow. The study results of this paper provide a certain reference for the further research on English grammar error correction algorithm based on the classification model.

1. Introduction

English grammar error correction algorithm is an important task of natural language processing, which uses computer programming technology to automatically recognize and correct the grammar, spelling, word order, and punctuation errors contained in English text written by nonnative language learners [1]. Due to the huge number of language vocabulary, the complexity of grammar rules, the ambiguity of semantics, and the ambiguity of speech, only the use of computer programming technology to analyze and study these problems will help improve the ability of language processing and recognition and enable natural language processing machines becoming machines with human intelligence [2]. There currently are three methods of grammar

error correction: the first is a rule-based method, which formulates specific error correction rules to correct specific error types and thus largely depends on the quality of the rules and can only modify certain types of errors; the second is a statistics-based method, which uses related information such as words to obtain text features, model the language, and select appropriate statistical models to correct errors in the text; the third is a depth-based method, which uses word vector representation to construct a deep neural network to correct errors in text end to end regardless of specific error types [3]. Therefore, all of these require a more accurate and efficient algorithm to correct English grammar, which is the main topic of this paper [4].

Classification is the core problem in machine learning and data mining and there are many classification models,

such as neural networks, decision trees, genetic algorithms, and support vector machines; in recent years, information and structure that have been extracted from reliable data classifiers have gradually become a hot research topic [5]. For traditional classification methods, the classification results are very clear, but if this clear classification method is used to classify fuzzy objects, it will not be in line with reality. Most classification models are based on the assumption that the distribution of classes in the data stream is roughly balanced and the designer usually assumes that the number of samples contained in the training data set is roughly the same [6]. This basic assumption is applied to many real data streams and it mainly uses the rules provided by the above modules to complete the automatic grammar inspection and correction function. For any automata constructed according to certain rules, the automata defines a language, which is composed of all the sentences that the automata can recognize and grammar and automata express language from the perspective of equivalent generation and recognition [7]. With this classification algorithm and the continuous increase of training samples and the continuous progress of the learning process, the classification accuracy of the resulting classifier will continue to improve, thereby reducing time and saving storage space [8].

On the basis of summarizing and analyzing previous research works, this paper expounded the research status and significance of English grammar error correction algorithm, elaborated the development background, current status, and future challenges of classification model, introduced the methods and principles of feature extraction method and dynamic residual structure, constructed a basic model for English grammar error correction based on classification model, analyzed the classification model and translation model of English grammar error correction, proposed the English grammar error correction algorithm based on the classification mode, performed the analyses of the model architecture and model optimizer of the grammar error correction algorithm, and finally conducted a simulation experiment and its result analysis. The study results of this paper provide a certain reference for the further researches on English grammar error correction algorithm based on classification model. The detailed arrangement is arranged as follows: Section 2 introduces the methods and principles of feature extraction method and dynamic residual structure; Section 3 constructs a basic model for English grammar error correction based on classification model; Section 4 presents the English grammar error correction algorithm based on the classification mode; Section 5 conducts a simulation experiment and its result analysis; and Section 6 is conclusion.

2. Methods and Principles

2.1. Feature Extraction Method. The mathematical system can give each node a minimum value; that is, the damping coefficient is added to the previous algorithm formula, and finally, the calculation formula for the node value is as follows:

$$Q_i(x) = \sum_{i=1}^n \left[\frac{d_i^2}{1-r_i} + \frac{B_i(x)}{A_i(x)} \right], \quad (1)$$

where $Q_i(x)$ is the total number of nodes in the network; d_i is the set of adjacent nodes of node i ; r_i is the outdegree of node i ; n is the total number of nodes; $A_i(x)$ is the directed edge weight of node i ; and $B_i(x)$ is the set of nodes with directed edges pointing to node i .

The back-propagation needs to calculate the error by comparing the prediction result of the forward propagation with the label of the training sample. For example, through using the squared error loss function, for the multiclass recognition problem of j categories and m training samples, the final output error function of the network is expressed as follows:

$$W_j = \frac{a_j^2 - b_j^2}{2} - k \sum_{j=1}^m \left(\frac{a_j + b_j}{c} \right)^2, \quad (2)$$

where W_j is the mapping matrix; a_j is the j th dimension of the label corresponding to the m th sample; b_j is the j th output predicted by the network for the m th sample; k is the error propagated to the previous layer; and c is the previous layer output.

Assuming that the sample vector $X = (x_1, x_2, \dots, x_n)$, the correlation degree of the vector X can be defined as

$$E(X) = \frac{R(X)}{\prod_{i=1}^n R(X)}, \quad (3)$$

where $R(X)$ is the expectation of vector X ; obviously, the greater the value of $E(X)$, the greater the correlation of vector X . In particular, if $x_1, x_2, \dots, x_n$ are independent of each other, then $E(X) = 1$; therefore, $E(X)$ can determine the weight coefficient in the above formula, but $E(X)$ needs to be estimated by the sample.

According to different methods, each basic classifier $t_i(x)$ is given a weight $y_i(x)$, and finally, each basic classifier is combined to form an overall classifier $T_i(x)$ to predict and classify the data in n , and the value of $T_i(x)$ and the calculation formula are as follows:

$$T_i(x) = \int_{i=1}^n \frac{t_i(x)}{y_i(x)}. \quad (4)$$

When the data class label in n is unknown, the $T_i(x)$ ensemble classifier considers the class distribution of $n-1$ to be the closest to the class distribution of n .

When integrating the projection index value, it is required that the local projection points of the projection value $U(i)$ are as dense as possible, preferably condensed into several points, and the projection points are scattered as much as possible on the whole. Therefore, the projection index function can be expressed as

$$U(i) = \sum_{i=1}^n \left[\frac{\sqrt{(p_i - s_i)(g_i - s_i)}}{C(x) - D(x)} \right]^2, \quad (5)$$

where $U(i)$ is the average value of the sequence; $C(x)$ is the window radius of the local density; $D(x)$ is the distance between samples; p_i is a unit order function; s_i is the number of indexes of the sample; and g_i is a sequence normalized to the characteristic value of the indicator.

2.2. Dynamic Residual Structure. The sample set of each index value is supposed to be $\{x(i, j) | i = 1, 2, \dots, n; j = 1, 2, \dots, m\}$, where $x(i, j)$ is the j th sample of the i index value and n and p are the number of samples and indexes, respectively. In order to eliminate the dimension of each index value and unify the variation range of each index value, the following formula can be used for extreme value normalization:

$$Y_{ij} = \sum_{i=1}^n \sum_{j=1}^m \left(\frac{\alpha_{ij}}{\beta_{ij}} - \gamma_{ij} \frac{l_{ij} + m_{ij}}{l_{ij} - m_{ij}} \right), \quad (6)$$

where l_{ij} and m_{ij} are the maximum and minimum values of the index; α_{ij} is the normalized sequence of the index eigenvalues; β_{ij} is the standard deviation of the projection value; and γ_{ij} is the local density of the projection value.

The mapping process is the density distribution formed by unsupervised clustering, and then the mapping relationship is formed through linear combination. The pattern vector uses nonlinear radial basis functions to obtain the competing attribution of each cluster center, and the most common representation of the radial basis function is as follows:

$$V_i = e^{w_i^2} - e^{u_i^2}, \quad (7)$$

where V_i is the output value of the middle layer i ; e is the pattern vector; w_i is the pattern vector of the cluster center corresponding to the middle layer; and u_i is the distribution width of the cluster center points, which can generally be attributed to this group according to V_i and training samples.

Information entropy represents the uncertainty of statistical characteristic quantities, which is defined as

$$N(V) = \prod_{i=1}^n \frac{h(V_i - K_i)}{q_i}, \quad (8)$$

where N is the training data set; k is the behavior category $\{K = 1, 2, \dots, K\}$; q_i is the proportion of the i th type of behavior data in the training data set N ; n is the number of thresholds; and h is the step size.

The adjacency matrix is used as the input of the algorithm, and the adjacency matrix vector of n nodes is used as the n -dimensional feature of the node. According to the multiple distances between the node and the i cluster centers x_i , the final division result is obtained as follows:

$$f(x_i) = \sqrt{\sum_{i=1}^n \frac{x_i^2 - z_i^2}{(n-1)(x_i + t_i)}}, \quad (9)$$

where $f(x_i)$ is the distance of point x_i ; z_i is the internal degree of node i ; and t_i is the dot product of the output of each module and the weight.

In the preprocessing stage, all are converted to lowercase, and the training set and test set are retained without stemming; the feature extraction algorithm uses the algorithm, and the word representation uses weights:

$$G_i(x) = \log[f_i(x) - o_i(x) - h_i(x)], \quad (10)$$

where $G_i(x)$ is the frequency of word x appearing in document i ; $f_i(x)$ is the total number of documents in the training set; $o_i(x)$ is the total number of words in the training set; and $h_i(x)$ is the document where word i appears number.

3. English Grammar Error Correction Model Based on the Classification Model

3.1. Classification Model. In the extraction part of the rule base, the rule base is combined with the corpus to filter out the rules of grammar errors and store them in the rule base to provide rule supplements for the subsequent automatic grammar checking and correction process. The grammar error correction model facilitates and quickly finds a list of documents containing words by establishing a mapping relationship between words and documents and inverting indexes are established offline in advance [9]. At present, the index object of the inverted index not only includes the words in the document, but also usually builds an index on the document itself. Therefore, not only the position of these words in a document, but also the storage location of the document itself is stored. When indexing words and documents, the words contained in the document will be automatically extracted to form a word dictionary, and the position of the word in the document will be placed in a fixed-format inverted file. In the inspection and correction part, it mainly uses the rules provided by the above modules to use the corpus, combined with the limited back-off algorithm, to complete the automatic grammar inspection and correction function. This part is the core function of the system and the system also improves the system through the cycle inspection module performance. In the process of automatic grammar checking, the scope of errors is determined by using the different collocation relationships of dependency analysis results (Figure 1).

In fact, the global model of distributed and streaming big data should be a generalization of the local flow data distributed in multiple nodes, and the main task is to be able to classify and predict the flow of unknown types of data, which is a distributed multiple node's streaming data providing a shared prediction model. In the global classifier, the basic operation operators involve the storage of the local mode, the generation of global training sample data, and the update and maintenance of the global integrated classifier. That is, through integrated learning technology, when the local patterns of all nodes in a history window reach the central node, the global pattern can be incrementally updated by triggering the global classifier. In theory, the reconstructed training sample set and the original data set need to be equivalent, and at least important statistical parameter values should be maintained. In the local classifier, the basic operation operators include

FIGURE 1: Classification model for English grammar error correction algorithm.

the acquisition of the current data block, the mining of the local pattern, and the updating and maintenance of the local pattern. According to the block-to-block mining technology, the incremental update of the local mode can be realized in a new data block arrival by calling the local classifier. The purpose of this is to ensure the difference between weak classifiers as much as possible to improve the coverage of the sample and also to improve the learning efficiency of the integrated classifier [10].

Validity means that when a rule is added to the rule base that can cause the wrong sentence to be detected, then this rule is valid. To achieve the effectiveness of the test, the mode can use as many descriptions as possible to express the scene used by the rule according to the situation and then evaluate the test effect. The difference from validity verification is that misjudgment verification estimates more things; misjudgment verification not only checks whether errors are ignored, but also judges the correctness of the errors pointed out. They use the characteristics of each grammar level to clarify the order and then store the rules according to the order. The reason for this method is that the system is arranged in the order of the frame of the file name, so the system intelligently reads one rule at a time and multiple-rule base in the layer rule model is composed of multiple rules, and one section stores a rule accordingly, and it is also a readable file with the rule base index table and the system depending on each other. The system will retrieve the index table according to the order of the index items before and after the index table. When the search reaches the end, it will immediately determine that the index table is empty; at the same time, the level numbers of both are expressed in the algorithm's input parameters.

3.2. Translation Model. There is a direct relationship between the core words in the main sentence. They should be put together to form a condensed clause, and the modifiers of each core vocabulary should also be put together with this core vocabulary as the center to form a condensed subclause sentence. If there are redundant components, they are placed on the next level, recursively forming a hierarchical structure. The vocabulary put together according to the grammar relationship forms a new clause, the words within the clause are mutually constrained, and the lower clause is the supplement and modification part of the upper layer. The use of different weights takes into account that different clauses have different influences on the entire sentence. For example, the main part of the whole sentence should contribute more to the correctness of the whole sentence than the main modification component. Among them, the error detection and correction module for general grammar errors provides alternatives for each word and then uses the language model to select the best choice for grammar error correction (Figure 2). The core of grammar error correction is to use the language model to decode the sentence to be tested. The correction of English gerund collocation errors is another application scenario for testing the effects of hierarchical language models. The gerund collocation is a kind of dependency, so there are certain advantages in applying this language model to the grammar error correction task of gerund collocation.

The output result of fuzzy comprehensive evaluation is no longer that an object clearly belongs to a certain category or does not belong to a certain category, but belongs to each category with a certain degree of membership. Such results are often more real and have more information, and

FIGURE 2: Translation model English grammar error correction algorithm.

sometimes, the actual application requires a clear classification judgment. In this case, it can be classified according to the membership degree of the object relative to each type and the deep model is essentially a deep mapping network structure. The input signal is continuously decomposed and expressed through layer-by-layer mapping in the network and finally forms a multilayer expression about the target, whose main feature is that it does not need to be artificially selected and displayed, but through automatic machine learning, gets a deep representation of the target. Some features that contribute a lot to the classification of rare categories will become unimportant if they are considered in the global context; the words that appear in the rare category are likely to be unusual words in the document set and are likely to be filtered out [11]. The classification accuracy of a particular category will be affected. In the first stage, increasing the feature dimension will add more classification information, so it will improve the classification performance; in the second stage, too many features will increase the classification information and also bring the noise and the classification will become worse.

In a typical generative confrontation network, the default generator models a mapping in a continuous space. However, in the grammar error correction task, the input and output of the generator are natural language sentences, which are essentially a sequence of discrete symbols. The typical generative confrontation network cannot be directly applied to the mapping modeling in the discrete space because it is impossible for the optimization algorithm to make minor updates to the generator parameters as in continuous space. The entire generative confrontation network is composed of a generator and a discriminator and the generator is a neural grammar error correction model, which is an encoder and decoder model that uses a sequence-to-sequence framework. The discriminator is a two-dimensional convolutional neural network and such a training strategy is fragile [12]. The classification model is responsible for distinguishing whether a corrected sentence is manually annotated or generated by a generator. Once the generator accidentally fails on some samples, so that the discriminator almost perfectly distinguishes the corrected sentence generated by it from the corrected sentence

manually labeled, then the generator is immediately lost. According to the reward value feedback from the discriminator, the generator knows that it has generated a poor-quality sequence, but the problem is that it does not know the correct direction to update the parameters.

4. English Grammar Error Correction Algorithm Based on the Classification Model

4.1. Model Architecture. Misjudgment verification, on the one hand, judges whether there is an error in the rule judgment and is ignored; on the other hand, it also judges whether there is a rule that fails to correctly indicate the error. Misjudgment is a problem often encountered in rule models, which is difficult to conflict between rules. The complete solution cannot be based on the rule model itself, because such consequences often lead to interlocking rules and mutual repulsion to make them too complicated. After the rule model detects the error, the model verifies the change of the sentence probability before and after the modification and determines whether to modify it. Similarly, after this, if the system finds that the modified error is an article or preposition error, the system calls this subsystem to judge the above sentence again. If the judgment is consistent, it is considered credible, as the previous experiments have shown that both the article correction system and the preposition correction system have a good recall rate. If the right and wrong can be verified before the grammar error correction, it is easier to improve the grammar error correction capability of the system (Figure 3). The technology uses random values to initialize the parameters of the model and then iteratively optimizes these parameters to make them closer to the optimal solution. These iterative optimization techniques ensure that every time they participate in optimization, they will be closer to the optimal value, but they may not be able to provide a method to determine whether the optimal value has been reached [13].

After analyzing a sentence through the dependency syntax analyzer, what is obtained is a set of dependent word pairs that represent the interdependence between the words in the sentence, and this set is also the set of semantic relations between words. The dependency relationship

FIGURE 3: Flowchart of English grammar error correction algorithm based on classification model.

between words in a sentence widely exists in the sentence structure of the text. The semantic relationship between words obtained after the dependency syntactic analysis of the sentence can well reflect the contextual structure information between the words in the sentence. A set of semantic relations between words can be obtained by performing dependency syntactic analysis on all sentences in a text, which can reflect the structural information of the text. The model uses the dependency relationship between feature words obtained after word segmentation of all sentences in the text and dependency syntax analysis to construct a text network. The text network constructed by this method has certain semantic connections between adjacent feature word nodes. Moreover, the text network constructed according to this method can not only reflect the contextual information of feature words, but also avoid the introduction of unrelated edges. In addition, using the dependency relationship between the feature words in the dependency syntax can realize the directed connection between the feature word nodes, thereby constructing the text network into a more accurate weighted directed network.

In the process of compound sentence modeling, according to the number of related words, the compound sentence modeling is divided into double related word compound sentence modeling, single related word compound sentence modeling, and unrelated word compound sentence modeling. For the modeling of double related words or single related word compound sentences, only the conditional probability of the corresponding related words needs to be substituted; for the case of unrelated words, since it involves the semantic research that is currently difficult to break through, here is a simple division of commas symbol. Prepositions indicate the relationship between nouns, pronouns, etc., and other words in the sentence and cannot be used as sentence components in the sentence alone [14].

The preposition and its object constitute a preposition phrase, which is used as an adverbial, predicative, complement, attributive, or prepositional object in a sentence. In future research, the model can study the grammar error correction of prepositions from the perspective of sentence composition and even sentence semantics. Considering automatic grammar error correction for missing errors, such as missing articles, missing prepositions, etc., it is difficult to accurately complete the task. Adding the wrong instance library is the solution to this problem, but the size of the instance library will reduce efficiency. Such methods can achieve high performance under the premise of sufficient annotation data, but there may be problems such as data sparseness and space overhead that affect the performance of grammar error correction, and statistical models require a large number of corpora to train.

4.2. Model Optimizer. Before the training starts, use the corpus preprocessing script to regularize the corpus and then perform model initialization and training. When an exception occurs in any step of corpus preprocessing, model initialization, and model training, a training exception notification will be sent to the administrator for system repair. After the model training is over, according to the training results, use the evaluation script of the previous grammar error correction experiment part to evaluate the model. If the grammar error correction effect is improved, update the grammar error correction model of the grammar error correction module; otherwise, directly end. Similar to grammar error correction, the design of feedback suggestions is also carried out from two aspects [15]. One is the feedback filtering interface itself, which provides its workflow chart; the other is the call flow between modules, which is explained using sequence diagrams. Abstract and extract the system function modules and build them into independent remote services to facilitate other business system calls to add new business functions and a new remote service module (Figure 4). They effectively realize the decoupling between the system and the modules, and it is also very easy to expand the new business. This module is mainly for implementing service cluster deployment and shielding application servers, using proxy servers to accept requests, making services transparent and isolating the real servers in the internal private network, and improving network security.

When solving the error polynomial equation, the money search method is used to solve the error polynomial equation obtained in the second step, and these solutions of the error polynomial equation are the error bits, and the corresponding error bits are corrected to realize the grammar error correction function of decoding. The error polynomial equation is a polynomial equation used to characterize the position of the error. Each of its solutions is the position where the error occurs. Therefore, to realize the grammar error correction function, the error polynomial equation must be obtained first and this mode does not need to support other parallel modes, or even serial mode, like the companion computing [16]. At the same time, after the money search retrieves the wrong location, it will

FIGURE 4: Algorithm processes of English grammar error correction based on classification mode.

automatically correct the data at the wrong location. This algorithm aims to determine whether there is similarity or even homology between two or more sequences by comparison, which aligns the sequences into an alignment table and then compares the base information in each column in turn. Similarly, grammar error correction through common structural features and the accuracy of grammar error correction in this way are better, but for a large number of data sets, sequence alignment is a relatively complex problem, and when looking for similar structural features, some sequences will be discarded, so the possibility of data loss also exists in the result of grammar error correction.

Although this design makes full use of context information, which greatly improves the translation effect, the disadvantage of this approach is that no matter how long the previous context is and how much information it contains, it will eventually be compressed into a few hundred dimensional vector, which means that the longer the context length, the more information will be lost in the final state vector and that after the length of the input sentence increases, the final decoded translation result will be significantly worse. In the basic model, every English word has the same contribution to the mode, which seems unreasonable; that is to say, the model that does not introduce attention has no problem when the input sentence is relatively short, but if the input sentence is, it is relatively long. At this time, all the semantics are completely represented by an intermediate semantic vector. The information of the word itself has disappeared, which makes it conceivable that a lot of detailed information will be lost. This is an important reason why the attention model is introduced and its geometric meaning is to perform a weighted summation of the information of all positions in the input

sequence, so as to achieve the effect of dynamically obtaining the most relevant subsequence information from the input sequence at any time stamp of the output sequence, that is, the above attention mechanism in the article [17].

5. Simulation Experiment and Analysis

5.1. Simulation Design. This study uses a grammar error correction test set for all types of English grammar errors as the experimental object; the dataset contains 1568 sentences, including 32 types of grammar errors. Two authoritative English-speaking annotators independently mark grammar errors and provide correct corrections. The community discovery result that does not contain any prior information depends on the selection of the initial clustering center. When the network structure is complex, the accuracy of the division will be greatly reduced. Therefore, some semi-supervised community discovery algorithms have been proposed and semisupervised learning uses both labeled data and unlabeled data, which can greatly improve the efficiency of learning. However, existing semisupervised algorithms require a large amount of labeled data, cannot really change the structure of complex networks, and still have the disadvantage of low efficiency. The place where the tag data is added is outside the complete algorithm framework and a complete algorithm includes multiple iterations. This method of adding does not make full use of the iterative process at each step. Prior information intervenes in the division of communities by means of random labeling, which requires a large amount of prior information, but usually manual labeling of prior information is difficult to obtain and expensive. Even if a large amount of prior

information is added, it still cannot really interfere with the correct division of the network; that is, the randomly added label is not the most needed part of the network division.

When the corpus is applied to different tasks, sometimes it is necessary to preprocess the corpus, such as using natural language processing technology to tag the corpus. When the corpus is small, some complex annotations can be performed on the entire corpus, such as some morphological and semantic annotations. More information can be obtained when analyzing annotated corpora, which are widely used in computational linguistics, speech recognition, machine translation, and other fields. The model extracts relevant information for the cleaned web pages, such as the keywords and keyword positions contained in the page content, and then uses the relevance algorithm to calculate the relevance of each web page to each keyword in the page content and hyperlinks, and finally, it uses these relevant information builds a web index database [18]. Nouns and verbs are not as limited as articles and prepositions, but each noun or verb has a corresponding variation. Therefore, in this paper, for noun errors, verb forms, and subject-predicate consistency errors, a change table is established for them by artificial induction and expansion and the size of this change table is not fixed and will gradually become larger after manual expansion. Since there are fewer article errors in the training corpus and more correct sentences, the extracted positive and negative samples are extremely imbalanced, and the imbalance of the samples will have a greater impact on the results of the recognition model.

5.2. Result Analysis. In the training phase, the feature vector is multiplied by the classification matrix, and the category with the highest score is used as the classification result. When the perception is classified incorrectly, the weight corresponding to the correct category is increased, and the weight corresponding to the wrong category is reduced. When the perception is classified correctly, there is no need to modify the matrix. Finally, all the classification matrices in the training process are added and averaged to obtain the final weight matrix. Obviously, compared to the voting naive perception, the average perception only records an additional cumulative value at any time, so it will be relatively efficient in terms of memory usage and training speed [19]. Compared with the statistical classification model, this deep classification model automatically learns the context information of the target word by a neural network, so complex feature engineering is not required (Figure 5). However, when the deep context model processes the context information, it not only inputs the native words of the sentence, but also the part of speech corresponding to the word is also added to some error types, so that the entire model has a good grammar error correction ability. Confusion network is a weighted directed graph, which has the characteristic that each path from the start node to the end node passes through all other nodes; it uses a neural network to extract the context information of the target word and does not require complex artificial feature engineering and the previous classification model.

FIGURE 5: Relationship between grammar error correction accuracy and residual value and algorithm training time with different recall rates.

All operations in the feature space are implicitly implemented through the inner product kernel function in the original space. That is, any algorithm that only uses scalar inner products can implicitly operate in a high-dimensional feature space through a kernel function. Therefore, this idea can be used to implement general linear algorithms in the feature space, but it can achieve nonlinear algorithms relative to the original space. This will greatly improve the efficiency of learning algorithms, improve existing algorithms, and improve various recognition rates of pattern recognition tasks (Figure 6). For the classifier base, the framework is transparent to them and they are only responsible for training the corresponding sample set and getting an output during testing. As for how the data set is generated, how the output is combined is not for the classifier base [20]. Therefore, each classifier base can be weighted in the decoding output process, based on the structure risk and regularization, to define the loss of the weight vector, and an optimal weight coefficient can be obtained by optimizing this loss function. This method can use function techniques to achieve nonlinear grammar error correction. In the training phase, each classifier is not optimized in the direction of the least overall loss; that is, the loss of each classifier is inconsistent with the overall multiclass classification [21].

The grammar error check is based on the spell check as the research premise. After the input text passes through the spelling processing module, the spelling error of the word has been corrected. When the part of speech of the confusing words in the confusion set is the same, the context word feature is more effective than the lexical collocation feature in the disambiguation process of the true word. On the contrary, when the part of speech is different, the lexical collocation feature is more effective. In some specific situations, if these two types of features are used to disambiguate true words at the same time, the problem of feature coverage and conflict may occur (Figure 7). If the correlation between the feature and the head word is relatively high, it means that when the feature appears, the head word is more likely to

FIGURE 6: Recall rate and classification accuracy with different training sample size in decoder and encoder.

FIGURE 7: Relationship between classification correct rate and repeat times in coder, classifier, and optimizer.

appear [22]. Here, mutual information is used to measure the two correlations; after the feature is selected, the weight is assigned according to the contribution of the feature. In addition, if the current feature has a high correlation with other confusable words in the confusion set where the head word is located, it means that its validity for distinguishing confusable words is relatively low, so the correlation and discrimination validity are used here and determine the weight of the feature. The feature of this method is that it combines the advantages of statistical methods and rule-based methods and automatically corrects major grammar errors in English grammar without manual intervention.

6. Conclusions

This paper constructed a basic model for English grammar error correction based on classification model, analyzed the classification model and translation model of English grammar error correction, proposed the English grammar error correction algorithm based on the classification mode, performed the analyses of the model architecture and model optimizer of the grammar error correction algorithm, and finally conducted a simulation experiment and its result analysis. In the inspection and correction part, it mainly uses the rules provided by the above modules to use the corpus,

combined with the limited back-off algorithm, to complete the automatic grammar inspection and correction function. In the grammar error correction task, the input and output of the generator are natural language sentences, which are essentially a sequence of discrete symbols and also improve the system through the cycle inspection module performance. The typical generative confrontation network cannot be directly applied to the mapping modeling in the discrete space because the optimization algorithm cannot be in continuous space, and minor updates are made to the generator parameters. For the classifier base, the framework is transparent to them and they are only responsible for training the corresponding sample set and getting an output during testing. The study results show that, with the continuous increase of training samples and the continuous progress of learning process, the proposed English grammar error correction algorithm based on classification model will continue to increase its classification accuracy, further refine its recognition rules, and gradually improve correction efficiency, thereby reducing processing time, saving storage space, and streamlining processing flow. The study results of this paper provide a certain reference for the further researches on English grammar error correction algorithm based on classification model.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest reported in this paper.

References

- [1] X. Xu, "Exploration of English composition diagnosis system based on rule matching," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 13, no. 7, pp. 161–173, 2018.
- [2] K. Dong-Ho, "Native and non-native English teachers' beliefs about grammar instruction and error correction in Korean college contexts," *Korean Journal of English Language and Linguistics*, vol. 16, no. 2, pp. 266–284, 2016.
- [3] A. Latif and M. Muhammad, "Sources of L2 writing apprehension: a study of Egyptian university students," *Journal of Research in Reading*, vol. 38, no. 2, pp. 194–212, 2015.
- [4] M. Angelova and Y. Zhao, "Using an online collaborative project between American and Chinese students to develop ESL teaching skills, cross-cultural awareness and language skills," *Computer Assisted Language Learning*, vol. 29, no. 1, pp. 167–185, 2016.
- [5] H. H. Pajouh, R. Javidan, R. Khayami, A. Dehghantanha, and K.-K. R. Choo, "A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 7, no. 2, pp. 314–323, 2019.
- [6] M. Autili, L. Grunske, M. Lumpe, P. Pelliccione, and A. Tang, "Aligning qualitative, real-time, and probabilistic property specification patterns using a structured english grammar," *IEEE Transactions on Software Engineering*, vol. 41, no. 7, pp. 620–638, 2015.
- [7] N. Macedo, T. Jorge, and A. Cunha, "A feature-based classification of model repair approaches," *IEEE Transactions on Software Engineering*, vol. 43, no. 7, pp. 615–640, 2017.
- [8] Z. Deng, L. Cao, Y. Jiang, and S. Wang, "Minimax probability TSK fuzzy system classifier: a more transparent and highly interpretable classification model," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 4, pp. 813–826, 2015.
- [9] H. Guo and W. Wang, "An active learning-based SVM multi-class classification model," *Pattern Recognition*, vol. 48, no. 5, pp. 1577–1597, 2015.
- [10] C. Lindström, "Two tenses are better than one," *International Journal for Lesson and Learning Studies*, vol. 6, no. 2, pp. 124–134, 2017.
- [11] T. G. Spies, R. Lara-Alecio, F. Tong, B. J. Irby, T. Garza, and M. Huerta, "The effects of developing English language and literacy on Spanish reading comprehension," *The Journal of Educational Research*, vol. 111, no. 5, pp. 517–529, 2018.
- [12] M. Antonelli, P. Ducange, F. Marcelloni, and A. Segatori, "A novel associative classification model based on a fuzzy frequent pattern mining algorithm," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2086–2097, 2015.
- [13] P. B. Gámez, N. K. Lesaux, and A. A. Rizzo, "Narrative production skills of language minority learners and their english-only classmates in early adolescence," *Applied Psycholinguistics*, vol. 37, no. 4, pp. 933–961, 2016.
- [14] J. Yang, K. Hotta, Y. Higo, H. Igaki, and S. Kusumoto, "Classification model for code clones based on machine learning," *Empirical Software Engineering*, vol. 20, no. 4, pp. 1095–1125, 2015.
- [15] G. A. A. J. Alkubaisi, S. S. Kamaruddin, and H. Husni, "Stock market classification model using sentiment analysis on twitter based on hybrid naive Bayes classifiers," *Computer and Information Science*, vol. 11, no. 1, pp. 52–64, 2018.
- [16] A. Schossee, S. Voigt-Zimmermann, S. Kropf, and C. Arens, "Evaluation eines Klassifikationsmodells horizontaler Gefäßveränderungen der Stimmlippen," *Laryngo-Rhino-Otologie*, vol. 95, no. 4, pp. 245–250, 2015.
- [17] S. Bittrich, M. Kaden, C. Leberecht, F. Kaiser, T. Villmann, and D. Labudde, "Application of an interpretable classification model on early folding residues during protein folding," *Biodata Mining*, vol. 12, no. 1, pp. 1–16, 2019.
- [18] W. M. Pearce, C. Williams, and W. Steed, "Dialectal grammatical differences in oral narratives of school-aged indigenous australian children," *International Journal of Speech-Language Pathology*, vol. 17, no. 4, pp. 335–345, 2015.
- [19] I. Simonova, "Blended approach to learning and practising english grammar with technical and foreign language university students: comparative study," *Journal of Computing in Higher Education*, vol. 31, no. 2, pp. 249–272, 2019.
- [20] N. Kahani, M. Bagherzadeh, J. R. Cordy, J. Dingel, and D. Varró, "Survey and classification of model transformation tools," *Software & Systems Modeling*, vol. 18, no. 4, pp. 2361–2397, 2019.

- [21] S. Gray, A. A. O'Connell, M. A. Restrepo et al., "The dimensionality of oral language in kindergarten Spanish-English dual language learners," *Journal of Speech, Language, and Hearing Research*, vol. 61, no. 11, pp. 2779–2795, 2018.
- [22] R. M. Nosofsky, C. A. Sanders, and M. A. McDaniel, "A formal psychological model of classification applied to natural-science category learning," *Current Directions in Psychological Science*, vol. 27, no. 2, pp. 129–135, 2018.