Journal of Network Intelligence ©2024 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **9**, Number **2**, May **2024**

Research on English Discourse Markers Based on Attention Mechanism AlexNet

Wen-Juan Ke*
School of Foreign Languages, Hubei University of Science and Technology
Xianning 437100, P. R. China
kewenjuancrystal@163.com

Zi-Jun Ji
Scuola Montessori Preschool, Tauranga 3110, New Zealand
420770416@qq.com

*Corresponding author: Wen-Juan Ke

Received November 6, 2023, revised January 1, 2024, accepted March 19, 2024.

Abstract. *For the purpose of dealing with the problem of low accuracy of English discourse marker recognition, this paper presents an English discourse marker recognition method on the ground of the attention mechanism AlexNet, which first uses the maximum posterior estimation to optimize the weight and bias parameters, and then uses the classified English discourse marker as the input parameter. Secondly, the high-dimensional sparse features are mapped to the low-dimensional dense feature space, and the input features are mapped to dense vectors. Then, the attentional mechanism is used to learn the relevance of the interactive behavior of various types of English discourse markers. At last, the prediction results were normalized by AlexNet and the final prediction results were obtained. Experimental results indicate that compared with the existing English discourse marker recognition methods, it can actually enhance the recognition performance of discourse markers, and obtain good outcome on the corpus, with an average recognition accuracy of 98.69%. It can be well be applicable for the study area of English discourse marker recognition.*

Keywords: english discourse; marker recognition; attention mechanism; AlexNet; feature space

1. **Introduction.** Discourse markers, which rely on the words after the preface and divide speech units [1], can restrict the listener's choice of contextual assumptions, and then infer the implied discourse intention. Discourse markers were discovered and explicitly mentioned as a distinct category in a lecture by Chaudron in 1986 [2]. He argues that these modifiers, you know, you see, well, have no effect on information from the point of view of grammatical structure, but are widely used in everyday speech. Since then, the recognition of the uniqueness of discourse markers in spoken language has gradually emerged. The 1950s and 1960s were just the beginning of the growth of linguistics, and the research of discourse markers did not get much attention. From the mid-1970s to the mid-1980s, with the gradual rise of discourse analysis, scholars have devoted themselves to the study of spoken language, and discourse markers have gradually received widespread attention [3,4,5]. Because of the existence of English language differences, traditional methods have difficulties in digital reading and low recognition accuracy when recognizing English discourse markers [6,7,8]. Therefore, the use of deep learning technology to build a recognition model of English discourse markers is of great research importance to

save manpower and material resources as much as possible in the recognition process and improve the recognition accuracy.

1.1. **Related Work.** In the study area of discourse markers, foreign scholars are mainly divided into two camps: one is the "coherence" school led by Redeker and Fraser, and the other is the "correlation" school led by Blakemore and Jucker. Redeker [9] believe that in Schiffrin's discourse coherence model, the two levels of information state and participation framework also exist in the other three levels and can be summarized into the other three levels. Fraser [10] believe that discourse markers include conjunctions, prepositional phrases, adverbial phrases, etc., which have their own syntactic characteristics and rich pragmatic functions in discourse. Jucker [11] believe that discourse markers play a significant role in discourse coherence and propose a discourse coherence model composed of five interrelated levels. Taguchi [12] suggest that discourse markers be regarded as a "pragmatic category", which mainly represents some connection between the current discourse and the previous discourse. Buysse [13] hold that before speaking, the speaker has formed in his mind the specific interpretation of views and the expectation that the listener can correctly understand them, and the listener must correctly process the speaker's words in order to accurately understand the speaker's intention. Asik and Cephe [14] believe that discourse markers in a discourse do not affect any semantic content, but express non-truth-valued semantic concerns, and help listeners to carry out cognitive reasoning in communication through the expression of procedural meaning of pragmatically restricted sentences. Crible et al. [15] divided discourse markers into receptive markers and exposative markers. Receptive markers, such as oh, yeah, okay, etc., are discourse markers indicating the listener's response to the information provided by the speaker; Declarative markers, such as, you know, well, etc., are speech markers used by speakers to provide information themselves. Cuenca and Crible [16] pointed out that from the perspective of syntax, discourse markers are not independent grammars, but mainly come from adverbial phrases, conjunctions and prepositional phrases. Dumlao and Wilang [17] analyzed and discussed the role of discourse markers in discourse relevance. In terms of the information of discourse context, discourse markers realize the coding process. Aiming at the problem that the extraction of character characteristic information in the above English discourse markers is not accurate, resulting in a low recognition accuracy, researchers have applied a series of deep learning algorithms, for instance, VGG [18], DenseNet [19], Res2Net [20], Vision Transformer [21] and Swin Transformer [22] have all been applied in the field of English discourse marker recognition. Becker et al. [23] proposed an English discourse marker recognition technology based on a small sample size. Deep convolutional neural network model was used for recognition, sliding window was used instead of manual segmentation, and the recognition accuracy reached 98.84%. Kumar et al. [24] improved the accuracy of English discourse marker recognition by using local and global attention mechanisms and convolutional neural networks to extract features of marker information for score prediction. Popescu-Belis and Zufferey [25] suggested an English discourse marker recognition algorithm based on the neural network of long-term memory. By optimizing and adjusting the model parameters, the average recognition accuracy at the letter level reached 97.69%. Borges et al. [26] combined the attention mechanism with the short-duration memory network to extract semantic relations, but the recognition rate was low. Because AlexNet network model can solve the overfitting problem, and can use multi-GPU acceleration calculation, it is also applied in the field of target classification recognition.

1.2. **Motivation and contribution.** Aiming at improving the recognition efficiency of English discourse markers, this paper proposes an English discourse marker recognition method based on the attention mechanism AlexNet. This method first optimizes the variable parameters of the AlexNet model, then preprocesses English discourse markers based on the improved AlexNet model, and then constructs their internal features by association. The similarity between the cross-features is calculated using multi-layer perceptron. The experimental results show that, contrasted with the present English discourse marker recognition methods, it can excellently enhance the recognition performance of discourse marker.

2. Theoretical analysis.

2.1. The definition of discourse markers.

Discourse markers are "non-independent units of discourse that connect discourse components", and they can play a role in local coherence for discourse [27]. Blackmore, a follower of Sperber and Wilson, believed that there was already a choice in the speaker's mind to interpret his words, and he wanted the listener to come to that choice as well. Therefore, the listener must handle the speaker's preset context correctly. Discourse markers are such a kind of language expression, which can minimize the cost of information processing by the listener and enable the listener to deduce the meaning that the speaker wants to express as soon as possible [28]. I mean, you see, in other words, that is to say, after all, anyway, however, nevertheless, actually, incidentally, etc., Some forms of expression in Chinese such as "but", "it is worth mentioning" and "frankly speaking", although they are part of the composition of discourse, they do not affect the truth condition of discourse or increase the propositional content of discourse, but only restrict the construction and understanding of discourse partially or as a whole. It is called a discourse marker [29]. Although linguists have different definitions of discourse markers, they basically agree that discourse markers have the following characteristics: They show a variety of categories, including words such as conjunctions (and, but), adverbs (actually, honestly), interjections (oh, well), phrases (in fact, by the way), and even short sentences (you know, I mean), etc. These terms have no influence on the truth condition of discourse, and express no conceptual meaning, only procedural meaning.

2.2. Attention mechanisms.

Attention mechanisms include Hard Attention, Soft Attention, Temporal Attention, Spatial Attention, and Convolutional Block Attention Module (CBAM) [30], this paper chooses the CBAM and adds convolutional neural networks to pay more attention to the target object in terms of channels and spatial dimensions, which has better explanatory properties.

Take the given text $G \in S^{D\times L\times V}$ as input. Two-channel attention is sequentially reasoned by channel attention mapping $N_D \in S^{D\times J\times 1}$ and spatial attention mapping $N_t \in S^{J\times L\times V}$, and the overall attention processing can be summarized as Equation (1) and Equation (2).

$$G^{*} = N_D(G) \otimes G \tag{1}$$

$$G^{**} = N_T(G^{*}) \otimes G^{*} \tag{2}$$

where $\otimes$ in the Equation (1) and Equation (2) represents the element-by-element multiplication operator.

Channel attention mapping is generated by the relationships between feature channels, treating each channel as a feature detector and therefore focusing more on the more meaningful parts of a given input image. For the purpose of calculating channel attention effectively, the spatial dimension of feature mapping is compressed and information is aggregated in the space. A common method is average-pooling. Max-pooling collects another important clue about the characteristics of different objects to infer more detailed channel attention. Therefore, both average-pooling and Max-pooling features are used in this paper. The shared network consists of a multi-layer perceptron (MLP) and a hidden layer, and generates spatial attention mapping by using the spatial relationship between features. The channel attention is calculated as shown in Equation (3).

$$\begin{aligned} N_D\left(G\right) &= \delta(MLP\left(\text{Avg}Pool\left(G\right)\right) + MLP(Maxpool(G))) \\ &= \delta(V_J(V_I(G^D_{avg})) + V_J(V_I(G^D_{\max}))) \end{aligned} \tag{3}$$

Different from the channel attention mechanism, the spatial attention mechanism pays more attention to the information part of the image input. This is complementary to the channel attention mechanism, and the spatial attention is calculated in Equation (4).

$$N_T\left(G\right) = \delta(g^{7\times 7}([\text{Avg}Pool(G); Maxpool(G)])) = \delta(g^{7\times 7}([G^T_{avg}; G^D_{\max}])) \tag{4}$$

The overall attention mechanism CBAM process is indicated in Figure 1.

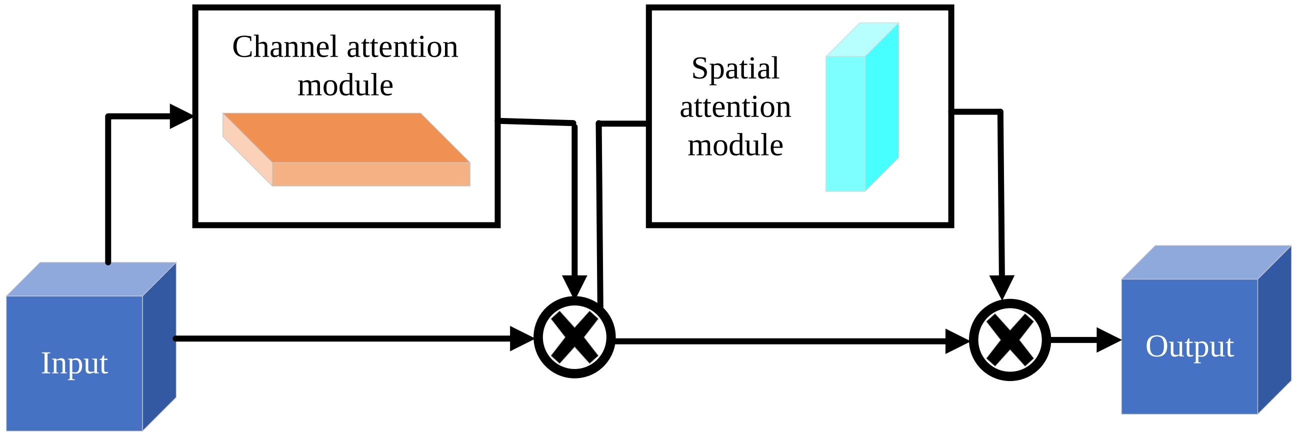

Figure 1. Flow chart of CBAM

3. Improved AlexNet network model.

3.1. Overall structure. The structure mainly makes up of four parts: input layer, multi-scale convolution layer, attention force layer and output layer. The model structure is indictaed in Figure 2.

In this paper, the multi-scale convolution layer is set as three CNN with one-dimensional convolution kernels of different sizes, and the convolution kernels are set as $c_1 = 2 \times 2$, $c_2 = 4 \times 4$, and $c_3 = 6 \times 6$ respectively, which are used to extract short, medium and long-term features of gas concentration sequences respectively. Specifically, in the DTH convolutional filter j of the first CNN layer, the activation value $\eta^f_{j,s}$ at time s is shown in Equation (5).

$$n^{(f)}_{j,s} = g\left(m^{(f)}_j + \sum_{s'=2}^{c} \left\langle V^{(f)}_{j,s'}, y^{(f-2)}_{s,c-s'} \right\rangle\right) \tag{5}$$

where c represents the convolution kernel size; g is the ReLu function. $\eta^{(f)}_{j,s}$ and $m^{(f)}_j$ are respectively the elements of $V^{(f)}$, $n^{(f)}$ and $m^{(f)}$.

After three layers of CNN, multi-scale features are fused in the feature fusion layer, and then standardized to obtain Equation (6).

$$x^{(f)} = \chi^{(f)} \frac{n^{(f)} - H[n^{(f)}]}{\sqrt{\mathrm{Var}[n^{(f)}]}} + \rho^{(f)} \tag{6}$$

where $H[n^{(f)}]$ is the average value of $n^{(f)}$, and $\sqrt{\mathrm{Var}[n^{(f)}]}$ is the standard difference.

In accordance with Equation (6), the active value matrix $x^{(f)}$ can be adjusted by $\chi^{(f)}$ and $\rho^{(f)}$. The last layer is the activation layer, which uses ReLu as the activation function to improve the convergence speed and robustness of the model. The final output of the multi-scale convolution layer is obtained from Equation (7).

$$X = g\left(f\left(d\left(n_t, n_n, n_f\right)\right)\right) \tag{7}$$

where n_s, n_m, and n_f are the extracted short, mediate, and long-term feature sequences respectively. $d(\cdot)$, $f(\cdot)$, and $g(\cdot)$ are the feature linkage operation, batch standardization operation, and activation value calculation respectively.

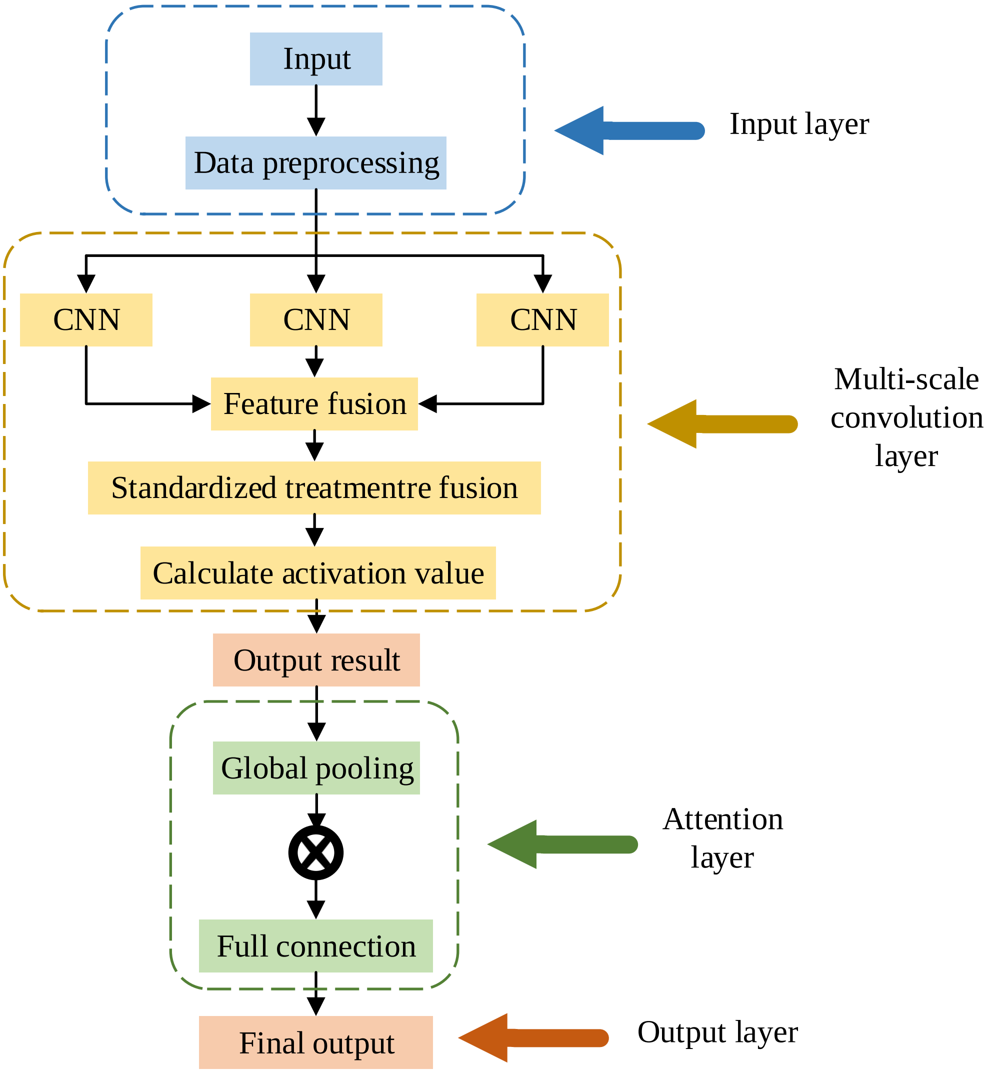

Figure 2. The Overall structure of the model

3.2. **AlexNet maximum a posteriori estimation optimization.** To optimize the weight and bias parameters of the AlexNet model, the variables F, G, and H are optimized using the maximum a posteriori estimation as follows.

$$\max_{F,G,H} p(F, G, H|S, Y, \delta^2, \delta_F^2, \delta_G^2, \delta_H^2) = \max_{F,G,H} \left[p(S|F, G, \delta^2) \cdot p(F|\delta_F^2) \cdot p(G|H, X, \delta_G^2) \cdot p(H|\delta_H^2)\right] \tag{8}$$

By taking the negative logarithm of both sides of the equal sign of Equation (8), it is reconstructed as the minimization loss function L.

$$L(F, G, H) = \sum_{j=1}^{M}\sum_{i=1}^{N}(S_{ji} - F_j^r G_i)^2 + \frac{v_F}{2}\sum_{j=1}^{M}\|F_j\|_2^2 + \frac{v_G}{2}\sum_{i=1}^{N}\|G_i - \mathrm{CNN}_H(Y_j)\|_2^2 + \frac{v_H}{2}\|H\|_2^2 \tag{9}$$

where v_F is $\frac{\delta^2}{\sigma_F^2}$, v_G is $\frac{\delta^2}{\delta_G^2}$, and v_H is $\frac{\delta^2}{\delta_H^2}$.

The coordinate descent method is used to minimize L, and the potential variables are iteratively optimized while the remaining variables are fixed. Equation (9) is a quadratic function about F. Meanwhile, G and H are temporarily assumed to be constant, and the loss function about F is considered. The function L can be obtained by differentiating F_j. The same operation for G results in the following representation.

$$f_j \leftarrow (GJ_jG^R + V_UJ_L)^{-1}GRS_j \tag{10}$$

$$G_i \leftarrow (FJ_iF^R + v_GJ_L)^{-1}(FS_i + v_HCNN_H(Y_i)) \tag{11}$$

where, J_j, J_i and J_L are the unit diagonal matrix, v_G and v_H are the equilibrium parameters. Equation (11) shows the effect of the markup latent vector CNN ($\mathrm{CNN}_H(Y_i)$) by v_H updated as a balance parameter g_i.

Then, the Equation (12) is obtained using the backpropagation algorithm commonly used in neural networks, and H is optimized until it converges or reaches a predetermined number of iterations.

$$\nabla_{h_l}\varphi\,(H) = -V_G\sum_{j}^{N} f\,(m_i)\,(G_i - \nabla_{f_l}CNN_H\,(Y_i)) + V_Hh_l \tag{12}$$

By optimizing F, G and H, we can finally achieve the recognition information of the predicted English discourse markers in Equation (13).

$$s_{ji} \approx E[s_j|f_j^RG_i, \delta^2] = f_j^Rg_i = f_j^S\,(CNN_H(Y_i) + \varphi_i) \tag{13}$$

4. Design of English discourse markers based on attention mechanism AlexNet.

4.1. **English discourse marker preprocessing.** On the ground of the AlexNet network model designed above, this paper uses classified English discourse markers as input parameters. These features cover the information of English teaching materials, grammatical difficulties, and synonyms of words, and the features include the ID, category, and discourse markers of teaching materials, as shown in Table 1. When preprocessing the input features, it is usually encoded as a one-hot vector, such as "[0, 1, 0]", so as to obtain a high-dimensional sparse vector representation. Considering that the increase of such dimensions and the sparse situation will easily lead to the risk of overfitting phenomenon, the sparse features are embedded into the low-dimensional and dense vector space, that is, the embedding layer. The specific operation of embedding is to multiply the $s \times t$ matrix composed of one-hot vectors with the embedded matrix of $s \times c$, and the result can be expressed as: $e = [e_1, e_2, \ldots, e_t]$. Where: c represents the number of features; $e_j \in S^{m\times c}$ represents the embedding vector of a feature; m is the original dimension of the feature; d is the embedded dimension, usually $c \ll n$, that is, the dimension of the embedded feature is much smaller than that of the original feature, which solves the problem of storage space waste caused by sparse data. These low-dimensional embedding vectors are then fed into shallow and deep models to be used as their input values.

Table 1. Classification of English discourse markers.

Textbook ID	**Category**	**Discourse marker**
1	Add class	And, also, too, in addition, moreover,...
2	Disjunctive class	But, yet, while, however, not...until,...
3	Selective class	Or, either...or, neither...or, nevertheless,...
4	Inference class	So, then, as far as I know, it is said that,...
5	Exegesis class	For example, and so on, such as, for instance,...
6	Enumerative class	Next, first, second, finally, after that, third,...
7	Destination class	So that, in order to, for the purpose of,...
8	Summary class	In the end, in a word, all in all, in short,...
9	Causation class	Because of, since, as, therefore, thus, due to,...
10	Result class	So, as a result, too...to, so...that, above all,...
11	Declarative class	That is to say, in other words, for my part,...

4.2. **Association construction of English discourse markers.** The association construction of English discourse markers can reflect the features between markers. In this paper, three dimensions of English textbook ID, category and discourse markers are selected for association construction. The three-level English discourse markers labeling architecture is shown in Figure 3.

The structures designed in this paper need to be simultaneously mapped into the same low-dimensional feature space, and each continuous feature corresponds to a feature vector in the embedding space. The AlexNet Product-based Neural Network (PNN) model based on the Embedding layer is introduced into the explicit second-order interaction layer after the embedding layer, which can well learn the special characteristics of high-dimensional sparse matrix. Firstly, the high-dimensional sparse features are mapped to the low-dimensional dense feature space, and the input features are mapped to dense vectors. Then, the attentional mechanism is used to study the correlation of interaction behaviors of various types of English discourse markers. Finally, the interaction features with attention weights are extracted and the implicit vector is obtained. Its loss function is expressed by Equation (14).

$$L\left(x, \hat{x}\right) = -x \log \hat{x} - \left(1 - x\right) \log\left(1 - \hat{x}\right) \tag{14}$$

4.3. **English discourse marker recognition based on attention mechanism AlexNet.** For the purpose of highlighting the features of English discourse markers, this paper calculates the weight of each feature through the attention mechanism, and updates the weight for the input vector sequence. The multi-layer perceptron has a multi-layer network structure, which can effectively extract high-order nonlinear features, so this paper uses multi-layer perceptron to calculate the similarity between the cross-features. The calculation equation is as follows.

$$S_n = \sum_{j=1}^{m} MLP(q_n, q_{1\cdots m}) \tag{15}$$

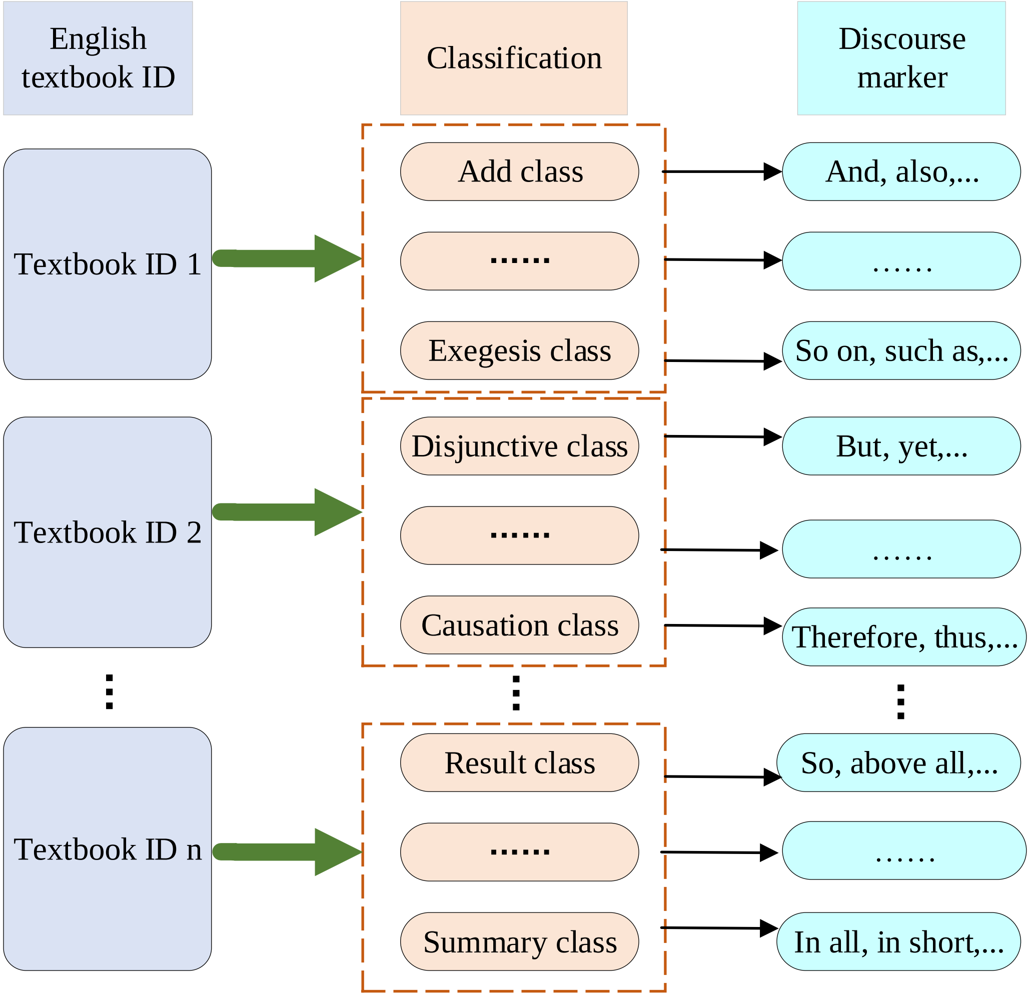

Figure 3. Three-level system structure

$$S_n = \sum_{j=1}^{m} MLP\left(r_n r_{1\cdots m}\right) \tag{16}$$

where: q_n represents the n-th interactive behavior feature vector in the feature matrix q; $q_{1\ldots m}$ represents all eigenvectors; S_n represents the n-th text feature vector in the feature matrix K; $r_{1\ldots m}$ represents all eigenvectors.

By normalizing the calculated similarity, the attention weight of each feature vector can be obtained by calculating $H_n = e^{Sim_{n_l}} / \sum_{j=1}^{m} e^{Sim_{n_l}}$ and $G_n = e^{Sim_{t_l}} / \sum_{j=1}^{m} e^{Sim_{t_j}}$ where h_n represents the attention weight coefficient of the n-th eigenvector in the behavior eigenvector matrix; g_n represents the attention weight coefficient of the n-th eigenvector in the tag feature matrix.The behavior feature weight coefficient and tag feature weight coefficient are represented by h_n and g_n respectively, and the attention weight coefficient of all feature vectors is splicing to get the behavior attention weight coefficient matrix $B(Q)$ and tag attention weight system matrix $B(S)$: $B(Q) = (h_1, h_2, \ldots, h_m)$, $B(S) = (g_1, g_2, \ldots, g_m)$.

By multiplying the behavior feature vector with the feature matrix Q' with weighted information and the tag feature vector with the feature matrix S' with weighted information, the final representation of the feature crossing is obtained: $G_{\text{att}} = [d_1, d_2, \ldots, d_j, \ldots, d_m]$.

Then AlexNet model can be used to fully model first-order features and second-order combined features, and the non-zero embedding vector corresponding to each feature can be obtained. The multi-layer perceptron can extract high-order nonlinear features to make up for the defect. Therefore, the AlexNet model is used in this paper to predict classification, and its output is expressed as Equation (17).

$$\hat{x}_{Alex}(y) = h_0 + \sum_{j=1}^{m} h_j y_j + \sum_{j=1}^{m} \sum_{i=j+1}^{m} h_{ji} y_j y_i \tag{17}$$

where h_0 represents the bias of the constant term, h_j is the weight of the j-th feature component, y_j represents the j-th feature component in the vector, y_i represents the i-th feature component in the vector, h_{ji} represents the interaction value between the j-th feature component and the i feature component, and m represents the number of features of the sample. The sum of the first two terms represents the linear part, and the third term represents the second-order feature crossing part. AlexNet model can calculate the similarity between two features without interactive data through inner product to ensure the completeness and consistency of feature learning.

The hidden layer output of CNN model is expressed as Equation (18).

$$\hat{x}_{CNN}(y) = f\left(h_l \cdots \left(f(h_1 y + a_1)\right) + \cdots + a_l\right) \tag{18}$$

where h_i represents the weight value of the i-th tag, and a_l represents the l-th layer bias.

The prediction results of the model were normalized by AlexNet, and the final prediction results were obtained. Normalization to Equation (19).

$$\hat{x} = f(\hat{x}_{\text{Alex}} + \rho \hat{x}_{\text{CNN}}) \tag{19}$$

where the nonlinear mapping function $f(\cdot)$ is the Sigmoid function and ρ is the compromise coefficient.

5. Performance testing and analysis.

5.1. Comparison of frequency and type of usage.

The experimental software environment is the integrated development platform python v3.6. The hardware environment is Intel(R) Core(TM) i5-10500 CPU @ 3.10GHz, 32GB memory, and 512MB storage space. In this section, TCIJ, the corpus of international journals, is used to evaluate the performance of this design method, and the method is contrasted with the method in literature [26]. Meanwhile, the influence of the search results of the frequency, type and location of English discourse markers on the experimental outcome is analyzed. For the convenience of description, the method designed in this paper is labeled OUR, and the method designed in literature [26] is labeled CSFD.

In this paper, the representative terms of each type of logical connectives in discourse markers are selected, and the listed representative search terms are searched in the corpus for frequency, and the standard frequency calculation and chi-square value test are carried out on the logical connectives in OUR and CSFD. The data results are shown in Table 2.

Table 2. Comparison of total frequency and chi-square value

Method	Characters' number	Frequency	Chi square value	P value
OUR	475639	7983	14.1529	0.000318
A	468423	7826	12.0628	0.002639

The frequency pairs of logical connection tags in OUR and CSFD are shown in the table above. Through searching, it is found that in the 475639 characters in the corpus, the logical link marker appears 7983 times; In CSFD's 468423 characters, the logical connection tag appears 7826 times, 157 times less than ours. In terms of the total frequency of use of logical connection tags, OUR is obviously higher than that of CSFD. In addition, from the calculation results, the chi-square value of the total frequency of OUR discourse markers is +14.1529, and the significance water-level P-value is less than 0.001, indicating that the frequency of use in OUR discourse markers is significantly higher than that of CSFD.

In this study, the representative search terms of each type of logical connectives were searched in the corpus for frequency, and the usage proportion of each type of logical connectives was calculated. The specific results are shown in Table 3 below.

Table 3. The frequency and proportion of various logical connection markers

Category	**OUR**		**CSFD**	
	Frequency	**Proportion(%)**	**Frequency**	**Proportion(%)**
Add Markers	613	7.68%	208	2.66%
Disjunctive Markers	658	8.24%	615	7.86%
Selective Markers	429	5.37%	367	4.69%
Inference Markers	715	8.96%	682	8.71%
Exegesis Markers	912	11.42%	961	12.28%
Enumerative Markers	106	1.33%	286	3.65%
Destination Markers	362	4.53%	981	12.54%
Summary Markers	1534	19.22%	1452	18.55%
Causation Markers	841	10.53%	796	10.17%
Result Markers	1294	16.21%	1008	12.88%
Declarative Markers	519	6.51%	470	6.01%
Total	7983	100%	7826	100%

On the ground of the frequency and proportion of usage, we can see that, first, the most frequently used logical connection Markers in OUR and CSFD are summative markers, followed by Result Markers. However, compared with CSFD, the two types of tags are used more frequently in OUR country than CSFD. Second, the discourse Markers with the lowest frequency are enumerative markers in OUR, and Add Markers in CSFD. In OUR approach, although Selective Markers were used infrequently, they were used significantly more frequently than CSFD. In addition, the frequency of Disjunctive Markers, Inference Markers, Causation Markers and Declarative Markers in our system is higher than that in CSFD.

5.2. **Performance analysis.** For the purpose of further evaluating the performance of the proposed method, a comparative experiment was set up. AlexNet model was used for data set training. In order to reduce the influence of some abnormal experiments, the whole experiment result is the average of 100 English discourse marker recognition results. Figure 4 shows the comparison of loss function values between OUR and CSFD models, and Table 4 shows the comparison results of each evaluation index.

It can be seen from Figure 4 that the loss function value of the proposed method reaches the optimal value of the 10 comparison models, and the loss value of the CSFD method is reduced by 0.2455. As can be seen from Table 4, among the four identification and evaluation indicators (accuracy, accuracy, recall, F1), the four indicators of the method in this paper reach the optimal level. The F1 value is shown in Equation (20).

$$F1 = [(Precision \times \mathrm{Re}\,call)/(Precision + \mathrm{Re}\,call)] \times 2 \tag{20}$$

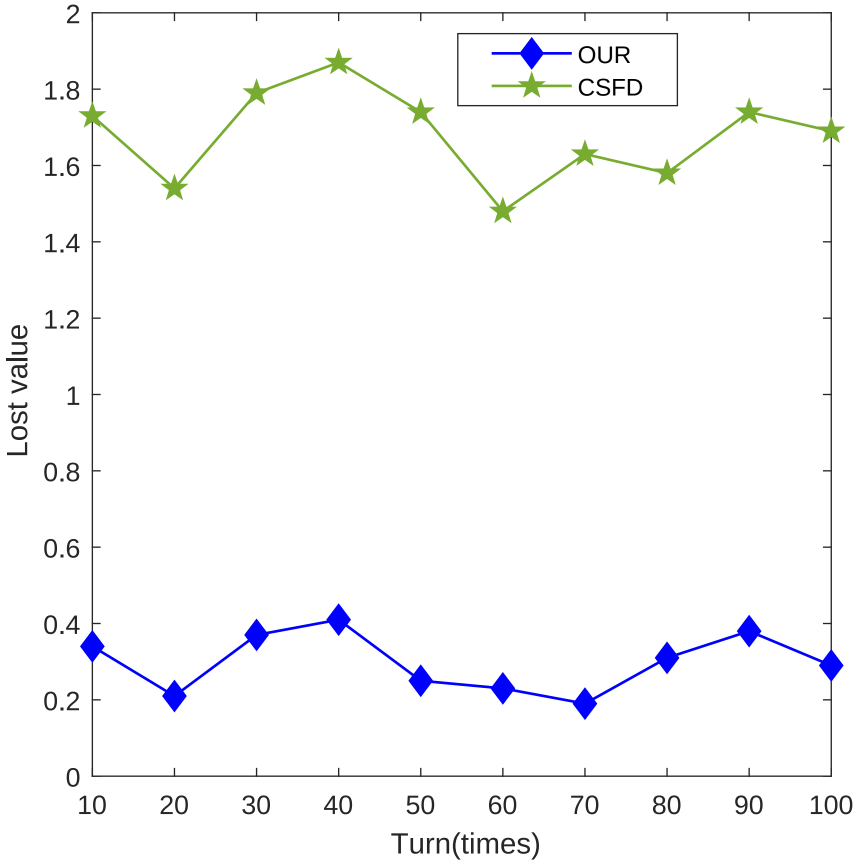

Figure 4. Comparison of lost value

Due to the improvement of ALexNet model in this paper, the accuracy rate, accuracy rate, recall rate and F1 are respectively improved by 3.44%, 3.26%, 1.83% and 2.55% compared with CSFD method, and various identification indicators of this method have been greatly improved. The results indicate that the enhanced model is effective in preventing overfitting and improving generalization. Contrasted with AlexNet network used in CSFD method, the model in this paper has high recognition accuracy, middle parameters, fast running speed, and the comprehensive performance is the best in the comparison model. Overall, the proposed method can achieve better simulation results in recognition of English discourse markers than other comparison models.

Table 4. Comparison of evaluation index results under different methods

Method	**Accuracy**	**Precision**	**Recall**	**F1**
OUR	99.58%	98.69%	99.15%	98.92%
CSFD	96.14%	95.43%	97.32%	96.37%

6. **Conclusion.** Aiming at improving the performance of English discourse markers, this paper designs an AlexNet English discourse marker recognition method based on attention mechanism. This method first fuses multi-scale features of the ALexNet model, then maps high-dimensional sparse features into low-dimensional dense feature Spaces, and maps input features into dense vectors. The interactive features with attention weights are extracted, and then the high-dimensional sparse features are mapped to the low-dimensional dense feature space, and the input features are mapped to dense vectors.

Then, the attentional force mechanism is used to learn the correlation of interaction behaviors of various types of English discourse markers, and finally, the similarity between cross-features is calculated using multi-layer perceptron. The experimental outcome indicates that compared with the present English discourse marker recognition methods, it can impactful enhance the accuracy rate, precision rate, recall rate and F1 of discourse marker.

Acknowledgment. This work supported by the Hubei Association of Higher Education in 2021, project number (No. 2021ZA22).

REFERENCES

[1] L. Fung and R. Carter, "Discourse markers and spoken English: Native and learner use in pedagogic settings," *Applied Linguistics*, vol. 28, no. 3, pp. 410-439, 2007.

[2] C. Chaudron and J. C. Richards, "The effect of discourse markers on the comprehension of lectures," *Applied Linguistics*, vol. 7, no. 2, pp. 113-127, 1986.

[3] Y. Ma, Y. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121-2133, 2022.

[4] F. Zhang, T.-Y. Wu, Y. Wang, R. Xiong, G. Ding, P. Mei, and L. Liu, "Application of Quantum Genetic Optimization of LVQ Neural Network in Smart City Traffic Network Prediction," *IEEE Access*, vol. 8, pp. 104555-104564, 2020.

[5] F. Zhang, T.-Y. Wu, J.-S. Pan, G. Ding, and Z. Li, "Human motion recognition based on SVM in VR art media interaction environment," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, 40, 2019.

[6] K. Wang, X. Zhang, F. Wang, T.-Y. Wu, and C.-M. Chen, "Multilayer Dense Attention Model for Image Caption," *IEEE Access*, vol. 7, pp. 66358-66368, 2019.

[7] K. Wang, C.-M. Chen, M. S. Hossain, G. Muhammad, S. Kumar, and S. Kumari, "Transfer reinforcement learning-based road object detection in next generation IoT domain," *Computer Networks*, vol. 193, 108078, 2021.

[8] W. Sun, "The Importance of Discourse Markers in English Learning and Teaching," *Theory & Practice in Language Studies*, vol. 3, no. 11, 2013.

[9] G. Redeker, "Linguistic markers of discourse structure," *Linguistics*, vol. 29, no. 6, pp. 1139-1172, 1991.

[10] B. Fraser, "An approach to discourse markers," *Journal of Pragmatics*, vol. 14, no. 3, pp. 383-398, 1990.

[11] A. H. Jucker, "The discourse marker well in the history of English1," *English Language & Linguistics*, vol. 1, no. 1, pp. 91-110, 1997.

[12] N. Taguchi, "A comparative analysis of discourse markers in English conversational registers," *Issues in Applied Linguistics-los Angeles*, vol. 13, no. 1, pp. 41-70, 2002.

[13] L. Buysse, "The business of pragmatics. The case of discourse markers in the speech of students of Business English and English Linguistics," *International Journal of Applied Linguistics*, vol. 161, no. 1, pp. 10-30, 2011.

[14] A. Asik and P. T. Cephe, "Discourse Markers and Spoken English: Nonnative Use in the Turkish EFL Setting," *English Language Teaching*, vol. 6, no. 12, pp. 144-155, 2013.

[15] L. Crible, L. Degand, and G. Gilquin, "The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis) fluency," *Languages in Contrast*, vol. 17, no. 1, pp. 69-95, 2017.

[16] M. J. Cuenca and L. Crible, "Co-occurrence of discourse markers in English: From juxtaposition to composition," *Journal of Pragmatics*, vol. 140, pp. 171-184, 2019.

[17] R. P. Dumlao and J. D. Wilang, "Variations in the use of discourse markers by L1 and L2 English users," *Indonesian Journal of Applied Linguistics*, vol. 9, no. 1, pp. 202-209, 2019.

[18] S. Tammina, "Transfer learning using vgg-16 with deep convolutional neural network for classifying images," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, no. 10, pp. 143-150, 2019.

[19] K. Zhang, Y. Guo, X. Wang, J. Yuan, and Q. Ding, "Multiple feature reweight densenet for image classification," *IEEE Access*, vol. 7, pp. 9872-9880, 2019.

[20] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, pp. 652-662, 2019.

[21] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, and Y. Xu, "A survey on vision transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 87-110, 2022.

[22] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang, "Ds-transunet: Dual swin transformer u-net for medical image segmentation," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-15, 2022.

[23] M. Becker, M. Bender, and M. Müller, "Classifying heuristic textual practices in academic discourse: A deep learning approach to pragmatics," *International Journal of Corpus Linguistics*, vol. 25, no. 4, pp. 426-460, 2020.

[24] A. Kumar, S. R. Sangwan, A. Arora, A. Nayyar, and M. Abdel-Basset, "Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network," *IEEE Access*, vol. 7, pp. 23319-23328, 2019.

[25] A. Popescu-Belis and S. Zufferey, "Automatic identification of discourse markers in dialogues: An in-depth study of like and well," *Computer Speech & Language*, vol. 25, no. 3, pp. 499-518, 2011.

[26] L. Borges, B. Martins, and P. Calado, "Combining similarity features and deep representation learning for stance detection in the context of checking fake news," *Journal of Data and Information Quality (JDIQ)*, vol. 11, no. 3, pp. 1-26, 2019.

[27] I. Vaskó, "Discourse Markers and Beyond–Descriptive and Critical Perspectives on Discourse-Pragmatic Devices Across Genres and Languages," *Orpheus Noster*, vol. 12, no. 1, pp. 118-120, 2020.

[28] B. Polat, "Investigating acquisition of discourse markers through a developmental learner corpus," *Journal of Pragmatics*, vol. 43, no. 15, pp. 3745-3756, 2011.

[29] M. R. Talebinejad and A. Namdar, "Discourse markers in high school English textbooks in Iran," *Theory and Practice in Language Studies*, vol. 1, no. 11, pp. 1590-1602, 2011.

[30] S.-H. Wang, S. L. Fernandes, Z. Zhu, and Y.-D. Zhang, "AVNC: attention-based VGG-style network for COVID-19 diagnosis by CBAM," *IEEE Sensors Journal*, vol. 22, no. 18, pp. 17431-17438, 2021.