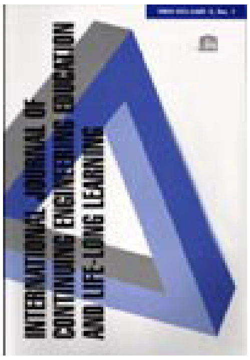
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Continuing Engineering Education and Life-Long Learning

ISSN online: 1741-5055 - ISSN print: 1560-4624
https://www.inderscience.com/ijceell

Evaluation method of Chinese grammar interactive teaching based on sentiment classification

Xue Wang, Guozhang Li

DOI: 10.1504/IJCEELL.2025.10074282

Article History:

Received:	27 February 2025
Last revised:	03 July 2025
Accepted:	12 September 2025
Published online:	28 November 2025

Evaluation method of Chinese grammar interactive teaching based on sentiment classification

Xue Wang

Public Teaching Department,
Hainan Vocational University of Science and Technology,
Haikou, 570100, Hainan, China
Email: 15203649991@163.com

Guozhang Li*

School of Information Engineering,
Hainan Vocational University of Science and Technology,
Haikou, 570100, Hainan, China
Email: 15203649997@163.com
*Corresponding author

Abstract: To better cultivate students' interest in learning Chinese grammar and master effective learning strategies, the interactive teaching method of Chinese grammar is widely used in the classroom. However, most students cannot be completely objective and fair when evaluating Chinese language interactive teaching, and cannot guarantee the objectivity and rationality of evaluation information. There are still many problems in the evaluation work. These problems also impact the original purpose of Chinese language teaching evaluation work, ultimately affecting the teaching efficiency and quality between teachers and students. This paper applies sentiment classification technology to the evaluation of Chinese grammar interactive teaching and conducts a brief research on the evaluation methods of Chinese grammar interactive teaching. It was proposed that the application of emotion classification technology in the current teaching evaluation system of colleges and universities would improve the evaluation quality of Chinese grammar interactive teaching by about 21.9%. Furthermore, it provided some impetus to the interaction between teachers and students, teaching quality, and teaching strategies of Chinese grammar.

Keywords: teaching evaluation; emotional classification; interactive teaching; Chinese grammar.

Reference to this paper should be made as follows: Wang, X. and Li, G. (2025) 'Evaluation method of Chinese grammar interactive teaching based on sentiment classification', *Int. J. Continuing Engineering Education and Life-Long Learning*, Vol. 35, No. 9, pp.72–89.

Biographical notes: Xue Wang is an Associate Professor in Hainan Vocational University of Science and Technology. She received her Bachelor's degree from Hainan Normal University, her Master's degree from Jose Rizal University. Her research interest focus is on Chinese language and literature.

Guozhang Li is an Associate Professor in Hainan Vocational University of Science and Technology. He received his Bachelor's degree from Nanchang University of Technology, his Master's degree from Jiangxi Normal University. He is studying for his PhD degree at University Malaysia Sabah (UMS). His research interest focus is on information technology (big data direction).

1 Introduction

Interactive teaching evaluation is one of the core links in the current Chinese grammar teaching. It not only helps to improve the efficiency and interactive quality of classroom teaching, but also provides important feedback information for teachers to improve their teaching strategies. However, in the actual teaching process, most students are often affected by subjective emotions, personal preferences, or temporary feelings when evaluating teachers' teaching behaviours, resulting in certain emotional tendencies and one-sidedness in the evaluation results, which makes it difficult to reflect the true picture of the actual teaching effect. This highly subjective evaluation method directly affects the objectivity and rationality of teaching evaluation data and reduces the pertinence and effectiveness of teaching improvement. To this end, this paper attempts to introduce sentiment classification technology into the interactive teaching evaluation system of Chinese grammar, and proposes a teaching evaluation method based on sentiment analysis, aiming to objectively and quantitatively analyse students' evaluation texts through natural language processing, identify the emotional expressions and attitude tendencies therein, and thus build a more scientific, transparent and fair teaching feedback mechanism. The ultimate goal is to realise a Chinese grammar teaching evaluation model for sentiment classification, enhance the information value of teaching evaluation, and promote the two-way improvement of teachers' teaching ability and students' learning motivation.

In this teaching evaluation system, data preprocessing plays a vital role as the basic link of text emotion recognition. In order to ensure the accuracy and semantic fidelity of corpus processing, this paper uses the Jieba Chinese word segmenter as the main text segmentation tool. It shows good compatibility and scalability in the educational context, especially supports embedding user-defined dictionaries, and can flexibly integrate common terms and sentiment lexicons in the teaching field. In terms of constructing the stop word list, we refer to the open source Chinese corpus of Tsinghua University, and further supplement special high-frequency stop words (such as '的'[1], '是'[2], '进行'[3], '方面'[4], etc.) according to the language characteristics of the teaching evaluation text to improve the efficiency of text processing and the accuracy of sentiment feature extraction. In addition, combined with the results of the part-of-speech filtering experiment, function words, punctuation marks, and neutral prepositions without emotional load are effectively eliminated, and standardised mapping rules for colloquial expressions are constructed. For example, '挺好' is simplified to '好'[5], and '不好懂' is uniformly treated as '难懂'[6]. In the synonym normalisation process, the method based on the semantic similarity of word vectors is used to merge words with similar expressions (such as '好懂', '易懂', '容易理解'[7], etc.) to enhance the consistency of feature dimensions and the generalisation ability of classification models. The entire preprocessing process is implemented modularly through Python scripts and relies on open-source platforms (such as GitHub) to share code and annotated data to ensure that data quality is controllable and the process is reproducible, providing stable and reliable data support for subsequent model training and deployment.

2 Related work

In recent years, Zhang (2017) scholars have applied emotion classification technology to many natural language tasks, but there is still a lack of corresponding research in the teaching of Chinese grammar interactive evaluation. Sentiment analysis in a broad sense is also called sentiment analysis and opinion mining. Sentiment analysis in this paper refers to understanding emotions such as joy, anger, sadness, and joy in the evaluation through interactive teaching evaluation texts. Wu (2022) proposed that opinion mining focuses on understanding the opinions and opinions expressed by students and teachers, that is, judging whether the comments express emotions or opinions. Peng (2019) believed that sentiment classification is mainly a process of analysing and reasoning on subjective comment texts with emotional color, that is, the attitude towards teaching comments is analysed, and whether they are positive or negative. Scholars such as Zhu and Huang (2021) have found that the sentiment classification of teaching evaluation is different from the traditional sentiment classification of text. The traditional sentiment classification is based on students' subjective judgments on knowledge, ability, and attitude in the learning process to obtain their value grades, and then, according to these results, to determine the scores given by teachers or give corresponding rewards. Such methods do not take into account the psychological changes of students. Liu and Xu (2020) take the sentiment analysis of teaching evaluation as a commonly used text sentiment analysis technique. It can not only be used to judge whether the text is positive, negative, or neutral, but it can also be used to understand the satisfaction of Chinese grammar interactive teaching. However, if sentiment analysis techniques are utilised, a well-labelled dataset is required, and a sentiment score is calculated using text classification to conduct an in-depth analysis of instructional evaluations. Therefore, this paper proposed a teaching evaluation system design framework based on deep learning and machine learning, which can automatically generate corresponding evaluation reports and provide the results to teachers and students in a structured form. In addition, the method has good practicality and can also effectively determine the sentiment classification of Chinese language teaching evaluation.

Teaching interaction and evaluation is the process of communication and negotiation between teachers and students, and is an indispensable and important link in teacher education and teaching activities. In the context of the new curriculum reform, Huang and Zhu (2021) believed that how to construct effective classroom teaching activities to stimulate students' interest in learning and improve classroom efficiency through research has become a hot issue for educators to research and explore. Through the analysis of Chinese language interactive teaching evaluation, scholars such as Shao-Chun L. learned that both sides of the interactive evaluation would have a profound impact on each other, which can not only improve information and thoughts but also produce emotional and psychological reactions (Lyu and Song, 2019). The interactive evaluation of both sides of teaching is based on the interaction between teachers and students, media, and resources. The realisation of the interactive teaching evaluation model requires the comprehensive use of multimedia, modern communication means, the Internet, professional laboratories, and corresponding teaching software. At the same time, teachers should set appropriate grading standards and grading methods according to students' participation in teaching activities. Hong (2019) found that the application of an interactive teaching mode can help stimulate the learning interest and enthusiasm of teachers and students, and promote the improvement of teaching quality. Scholars such as Endo et al. (2017) believed that the interactive-based teaching evaluation should be carried out synchronously with the teaching, the evaluation content should be consistent with the teaching content, and the teaching objectives should be completed together. Therefore, the construction of teaching evaluation criteria should point to the activities of teaching and learning, and jointly promote the development of teaching and learning. Through research, Weurlander et al. (2017) found that Chinese-language interactive teaching evaluation can have a positive impact on improving teaching cognition, promoting knowledge transformation, classroom analysis and diagnosis, and classroom feedback regulation. Borkovskaya et al. (2018) believed that effective teaching evaluation needs to collect data through a corresponding system, and analyse the relevant information of interactive teaching. On this basis, a valuable judgment was proposed. Effective teaching evaluation criteria should be designed with purpose and pertinence, to form a real and effective standard scale for the quality of teaching and learning. Therefore, it is necessary to design corresponding evaluation standards for teaching evaluation under the interactive mode, to promote the improvement of classroom teaching effect and enhance the professional ability and professional quality of teachers.

According to the existing research, the interaction of Chinese grammar teaching based on emotion classification means that, under the emotion classification, learners and various Chinese grammar learning resources can have related effects and influence each other. In this way, a rich language situation was created, which stimulated the interest of scholars in learning Chinese grammar (Yakovleva and Yakovlev, 2018).

3 Sentiment classification technical evaluation process

3.1 Sentiment classification

The research on emotion classification technology is an important research field in the evaluation of interactive teaching of Chinese grammar, which is of great significance to the evaluation of interactive teaching. Text words in teaching evaluations do not all have similar emotional tendencies and intensities. Therefore, it is very important for the sentiment classification task to better encode the context and extract the key information in teaching evaluation from it (van Ewijk, 2021). In the past three years, some scholars have introduced the attention mechanism of text in NLP. Especially combined with the classic Encoder framework, SoftAM, GlobalAM, and Hierarchical AM series attention models were proposed, which have been successfully applied in many fields (Eiman and Matthew, 2017). Due to the difference in granularity in processing object texts, when researching text sentiment classification in teaching evaluation, it is generally carried out from multiple different levels, such as document level, paragraph level, and word level (Hamker, 2018).

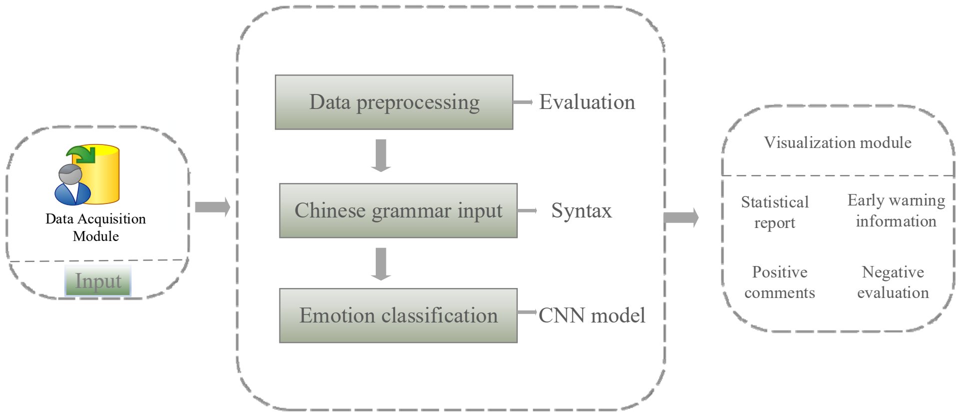
Sentiment classification technology can classify massive teaching evaluations through a special network model. The core idea of the technology is to use different granularity relationships between teaching evaluations to classify (Yu et al., 2018). Therefore, this study proposed a hierarchical attention network framework based on emotion scoring, in order to realise the effective classification of text emotion in the teaching evaluation process. A bidirectional recurrent neural network encoder is used to encode the word vector and sentence vector, respectively, in the teaching evaluation, and the final expression of the document is obtained through the weighted summation of the attention mechanism of the network model (Shermin et al., 2021). The neural network model also designed a sentiment scoring system for the words and sentences of the text in the auxiliary teaching evaluation, and used the scoring system to adjust the attention weight distribution (Ypsilantis and Montana, 2017). The biggest feature of teaching evaluation emotion classification is that it involves words with emotional tendency and a certain degree of discrimination. This is also an important difference between sentiment classification and topic classification in teaching evaluation, so how to extract effective sentences that can express emotion in teaching evaluation becomes the core point. After exploring the influence of the evaluation of emotional information on the classification performance, it is found that the subjective sentences of teaching evaluation contain most of the required emotional characteristics, so the subjective analysis of the evaluation input sentences becomes a key step.

This article adopts the emotional analysis technology method based on deep learning to analyse and study the emotional tendency information contained in the evaluation of Chinese grammar interactive teaching, and monitor the emotional text through the system. Sentiment classification analysis is carried out in five steps: acquisition of evaluation information, preprocessing of text data, word vectorisation, learning and analysis of evaluation text data, and visualisation of results. First, the system collection technology is used to complete the data collection (Kim and Hahn, 2018). Then, on the teaching evaluation data, the Chinese sentiment segmentation is performed to stop the word operation and save the key information of the sentence. Finally, the word embedding tool is used to convert words into word vectors that can be learned by a convolutional neural network (CNN) on the evaluation system. Through the learning of the system crawling process, the main features are extracted, and the classification of sentiment evaluation is finally completed (Eiman and Matthew, 2017).

3.2 Interactive Chinese grammar teaching evaluation design process

In this study, the CNN model is used to solve the emotional tendency analysis in interactive teaching evaluation. The word vector matrix obtained by the transformation of the teaching evaluation is used as the input of the CNN. The text features extracted by the convolution layer are used, and then the maximum pooling method is used to reduce the dimension of each evaluation feature vector. Finally, the ReLU function is used to classify the text in the fully connected layer and output the teaching evaluation. CNN is a multi-layer supervised learning neural network. The neural network consists of five sections, namely the input layer, the convolution layer, the pooling layer, the fully connected layer, and the output layer. The convolutional layer and the pooling layer constitute the main modules of the text feature extraction function. Figure 1 is a model of a CNN.

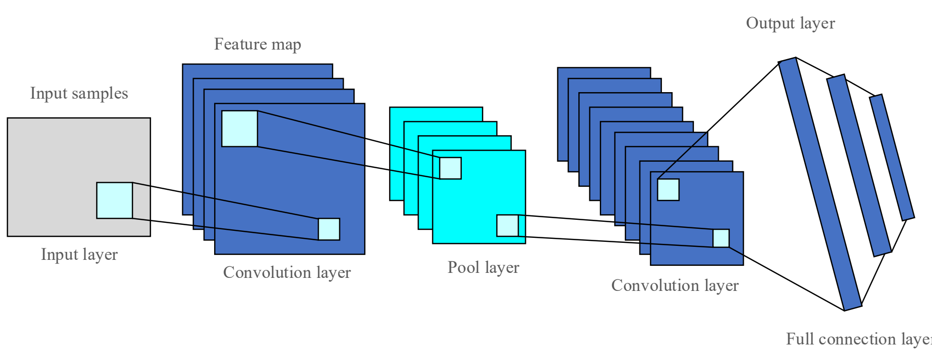
Figure 1 Flow chart of Chinese language teaching evaluation design system (see online version for colours)



3.3 Deep learning sentiment classification model

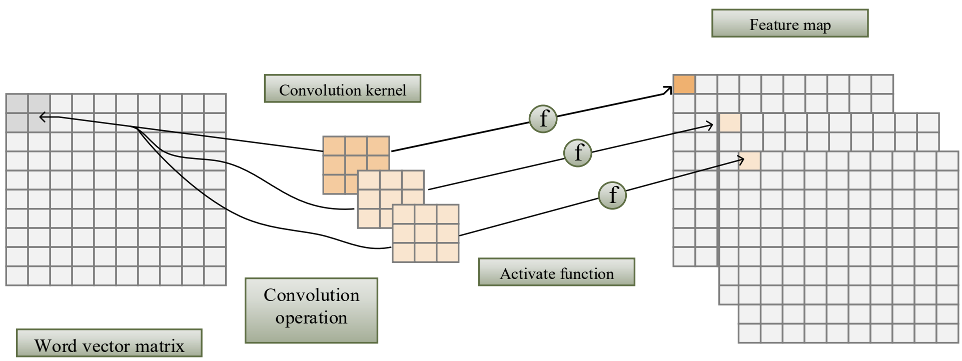
- *Convolutional layer:* the vector matrix of teaching evaluation words is input into the convolution operation of the first layer of the convolutional layer, and the corresponding emotional text feature map can be obtained. Through the convolution operation, the original signal can be enhanced, the noise can be reduced, and the input text can be convolved. Finally, different feature extractions are performed on the input samples. At the same time, there can be several different types of convolution kernels in each convolution layer, and each convolution kernel corresponds to a feature map. The sentiment classification vocabulary corresponding to each feature map is different, as shown in Figure 2.

Figure 2 CNN structure (see online version for colours)



- *Pooling layer:* after the convolution operation, the size of the feature matrix in the teaching evaluation is usually large. The pooling operation is used in this paper, which can not only reduce the dimension of the word vector matrix generated by the convolution layer, but also effectively reduce the number of parameters and the amount of calculation. In this experiment, the maximum pooling method is used to reduce the dimension of the word vector matrix. After the convolution operation, the teaching evaluation feature matrix is divided into multiple rectangular regions, and the maximum value of each sub-region is output, which reduces the size of the data space, as shown in Figure 3. When the feature matrix dimension reduction is used for teaching evaluation, the feature representation is more robust to changes in the position of the input word vector. At the same time, overfitting can be avoided to a certain extent.

Figure 3 Convolution layer with multiple convolution kernels (see online version for colours)



- *Fully connected layer:* the fully connected layer is a 'classifier' of the entire CNN model. The full connection maps the 'distributed feature representation' in the sample label space in a 'learning' way. In the sample label space, gray neurons represent that the sample feature has been found (activated), and the obtained activation value is obtained by the CNN model extracted from the teaching evaluation text information.

In response to the need for more fine-grained emotion recognition, this paper proposes an extension of the multi-label classification framework. By changing the output layer from binary classification (positive/negative) to multi-label (confused/satisfied/frustrated/neutral) and introducing a hierarchical label embedding strategy, the model can capture complex emotions (such as 'satisfied but confused') at the same time. Although the macro-average F-measure of this extended framework on the four-classification task is lower than the binary classification result, it provides a richer decision-making basis for emotional intervention in subsequent teaching scenarios.

4 Interactive evaluation algorithm of emotion classification technology

Given a sentence of input teaching evaluation, the AttLSTM model can be expressed as a function, assuming

$$\{y\} = y_1, y_2 \cdots, y_m \tag{1}$$

A set of polar categories is established for comment sentiment; that is, the purpose of the AttLSTM model is to calculate the likelihood probability in each category.

The sentence of x the input teaching evaluation is assumed to contain a total of n words, denoted as $x = x_1 * \ldots * x_i * \ldots * x_K$.

Among them $x_i(1 \leq i \leq k)$ is the i word in x, and $*$ it is the concatenation operation between words.

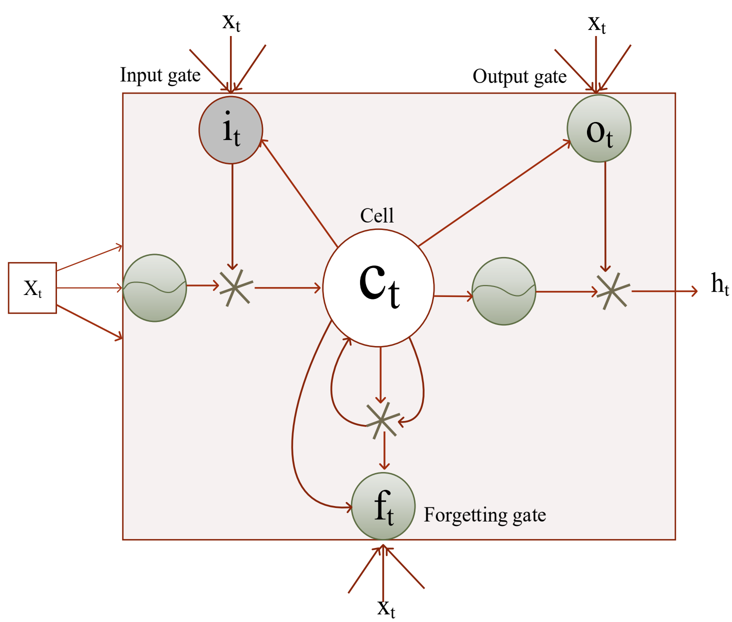
This article directly uses sentences representing teaching evaluation. In order to facilitate unified processing, this study compensates for sentence length in teaching evaluation. The sentence length threshold is assumed to intercept only the previous word for teaching evaluation texts longer than that, and use specific symbols (e.g., long PAD/short PAD) for teaching evaluation texts that are shorter than that length. The teaching evaluation text is taken as an example, and the corresponding completion operations are:

$$x_{1:k} = \begin{cases} x_1 * \cdots * x_n & k \geq n \\ x_1 * \cdots * x_n * \{\langle PAD/\rangle\} & k \geq n \end{cases} \tag{2}$$

Each word $x_i(1 \leq i \leq n)$ in the teaching evaluation text is mapped to a continuous, low-dimensional, dense real vector containing semantic emotion information.

Under the AttLSTM model, the forward LSTM network is used to encode the teaching evaluation text, and the long-distance dependencies between the vocabulary of the teaching evaluation text are learned. Its structure is shown in Figure 4.

Figure 4 Components of LSTM (see online version for colours)



Among them, $(1 \leq t \leq n)$ the current time c_t represents the hidden state of the neuron corresponding to the input value, h_{t-1} is the output value of the previous time point, o_t is the current input, x_t is the output, W_i, W_f, W_c, W_O, U_i, U_f, U_C, U_O and V_O are the corresponding weight matrices, b_i, b_c, b_f and b_o is the bias vector. It is usually sigmoid, and then the LSTM network calculation process is like this:

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{3}$$

$$c_t = \tanh\left(W_C x_t + U_c h_{t-1} + b_c\right) \tag{4}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{5}$$

$$o_t = i_t * c_t + f * c_{t-1} \tag{6}$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + V_o c_t + b_o\right) \tag{7}$$

$$h_t = o_t * c_t \tag{8}$$

The attention layer of the AttLSTM model mainly learns the weight value of each position, so that the more important words in the teaching evaluation text correspond to a higher attention value to highlight the effect of the important words on the evaluation results. First, through the self-attention model, the weight value of each word in the sentence is calculated to form the attention word vector. Then, the encoded output word matrix is weighted and summed to form a new semantic representation of the text. Specifically, for each input teaching evaluation text $x \in R^{n \times d}$ and the corresponding coding output $H \in R^{1 \times n \times r}$, the calculation formula of the attention word vector $\alpha = \{a_t\}(1 \le t \le n)$ and the final word order representation h of the text is as follows:

Given the target classification label $y = \{y_1, y_2, y_3, \ldots, y_m\}$, when $m = 2$, y is the classical binary sentiment classification (such as positive/negative), that is, the logistic regression based on sigmoid; when $m > 2$, y is the classification of multiple logistic sentiment regression. Specifically, it can be described as:

$$f : x \to y = \{y_1, y_2, \cdots, y_m\} \tag{9}$$

Assuming that the symbols represent all the parameters of the AttLSTM model, for a given $x\theta$, the output layer converts the calculation result into a conditional probability distribution $P(y \mid x, \theta)$ for each element in the set y.

The training set is given:

$$T = \left\{\left(x^{(i)}, t^{(i)}, y^{(i)}\right) \middle| 1 \le i \le |T|\right\} \tag{10}$$

Then the predicted probability distribution is:

$$f\left(x^{(i)}, \theta\right) = \frac{1}{\sum_{j=1}^{m} \exp\left(y_j \mid x^{(i)}, \theta\right)} \begin{bmatrix} \exp\left(P\left(y_1 \mid x^i, \theta\right)\right) \\ \exp\left(P\left(y_2 \mid x^i, \theta\right)\right) \\ \vdots \\ \exp\left(P\left(y_m \mid x^i, \theta\right)\right) \end{bmatrix} \tag{11}$$

The first paragraph to the right of the equal sign is the normalised result of the probability distribution, and the median value of the m-dimensional vector is the maximum value of $\max(f(x^{(i)}, \theta))$. The corresponding label is the final prediction result of the model, namely:

$$y^{(i)} = \max\left(f\left(x^{(i)}, \theta\right)\right) \tag{12}$$

5 Teaching evaluation model training

The AttLSTM model uses cross-entropy as a loss function for sentiment classification. For the training set:

$$T = \left\{ \left(x^{(i)}, t^{(i)}, y^{(i)} \right) \middle| 1 \le i \le |T| \right\} \tag{13}$$

For any teaching evaluation text *x*, its softmax loss function is defined as follows:

$$J\left(x^{(i)}, \theta\right) = \begin{cases} -t^{(i)} Iny^{(i)} - \left(1 - t^{(i)}\right) In\left(1 - y^{(i)}\right), m = 2 \\ -\sum_{i=1}^{m} \left(t^{(i)} Iny^{(i)} + \left(1 - t^{(i)}\right) In\left(1 - y^{(i)}\right)\right), m > 2 \end{cases} \tag{14}$$

The loss function $L(T, \theta)$ of the entire teaching evaluation training set *T* is defined as:

$$L(T, \theta) = \frac{1}{|T|} \sum_{i=1}^{|T|} J\left(x^{(i)}, \theta\right) \tag{15}$$

Based on the loss function, the model optimises the loss problem by iteratively solving the loss value and gradient descent, so that the value of the loss function is minimised. To improve efficiency, the model uses the miniGbatch method to set the loss value of each batch to miniGbatch, and the number (batch-size) *K* is usually much lower than that. At this time, the loss function is:

$$L\left(x^{(i:i+K)}, \theta\right) = \frac{1}{K} \sum_{j=i}^{i+K} J\left(x^{(j)}, \theta\right) \tag{16}$$

AttLSTM uses the RMSProp-based optimiser. In order to avoid the problem of over-fitting, this paper applies the dropout forwarding strategy to the input layer and encoding layer of the LSTM network, respectively. The corresponding dropout value and other key parameters are listed in Table 1.

Table 1 Key parameters of AttLSTM

Layer	*Parameter*	*Value*
Input layer	The vector dimension of words	300
Coding layer	Input dimension	128
Attentional layer	Weight dimensions	128
Input layer	Number of hidden neurons	128/relu
	Activation function	
Trainer	classifier	Rmsprop
	Loss function	Multivariate crossover

The selection of model parameters in this paper is based on dual verification of ablation experiments and literature references. The word vector dimension (300D) refers to the optimal configuration of Word2Vec in Chinese text, and the classification performance under 100D, 200D, and 300D is compared through ablation experiments. It is found that the F-measure is improved by 1.2% at 300D. The learning rate (RMSProp optimiser initial value 0.001) is determined by grid search. In the range of 0.0001 to 0.01, this value makes the model converge fastest and minimise overfitting on the validation set. In addition, the design of the encoding layer input dimension (128) and the attention weight size (128) follows the 'feature compression-restoration' balance principle to avoid the loss of semantic information due to insufficient dimension.

6 Experimental results of emotional words

6.1 Emotion recognition

In this study, multiple loops are adopted, and a step-by-step screening strategy is adopted. The emotional word set and other existing emotional dictionaries are used as the candidate emotional word A table in this study, and the candidate emotional word A table is combined with the existing emotional dictionary. All the words in the candidate emotional word list A are assumed to be emotional words. Through multi-feature linear fusion, the emotional tendencies of all words in Table A are calculated, and a new emotional word table B is constructed on this basis; the initial state of B is an empty set. If the emotional word table A is inconsistent with the emotional word table B, it means that the candidate emotional word table B selects non-universal emotional words. Next, A is replaced with B, and the sentiment tendency of the words in the sentiment vocabulary A is recalculated until the candidate sentiment vocabulary A is consistent with the new general sentiment vocabulary B. So far, the new B-list has become an emotional word for teaching evaluation.

The specific implementation steps are:

Step 1 The existing sentiment dictionary is used to construct the candidate sentiment word list (A) in this paper, and construct the blank sentiment word list (B).

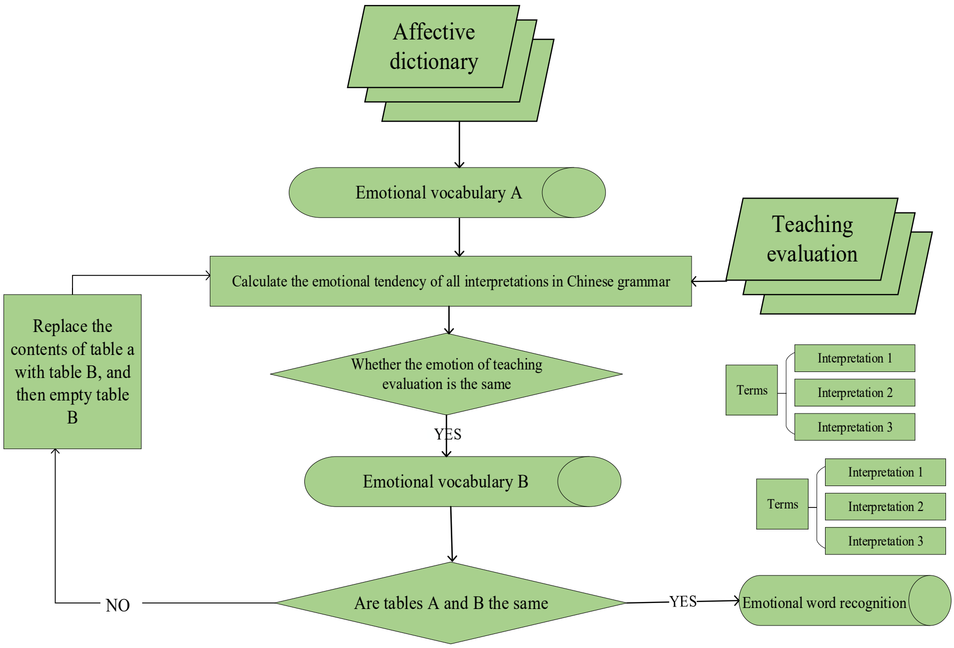
Step 2 For each word in the candidate sentiment word list A, the sentiment classification method in the text is adopted, and the sentiment tendency of all their explanations is calculated. However, the words with all the same interpretations are added to Table B as emotional words in the text.

Step 3 The word A in the candidate's emotional vocabulary is compared with the B word in the general emotional vocabulary. If it is consistent, then step 4 is carried out. Otherwise, word A is replaced with word B, and the process is skipped to step 2 to continue.

Step 4 The newly generated emotional vocabulary B is the final result. The algorithm process of constructing the emotional dictionary in this paper is shown in Figure 5:

In order to solve the problem of matching general sentiment dictionaries with teaching scenarios, this paper introduces a domain adaptation strategy in the dictionary construction process. Specifically, in the stage of screening candidate sentiment words, the intersection ratio of high-frequency words in the teaching evaluation corpus and the general dictionary (such as the coverage of terms such as 'classroom interaction' and 'grammatical difficulties') is statistically analysed, and the weight of sentiment words is adjusted by combining the term frequency-inverse document frequency (TF-IDF) weighting. In addition, for domain-specific vocabulary not covered in the general dictionary (such as 'teaching rhythm' and 'exercise design'), it is supplemented to the dictionary through manual annotation and iterative optimisation.

Figure 5 Emotion recognition process (see online version for colours)



6.2 Experimental data

To test the feasibility of the sentiment word recognition method proposed in this paper, experiments were carried out using four existing sentiment word databases. Based on the 'Modern Chinese Dictionary', the words appearing in the 'Modern Chinese Dictionary' were extracted as candidate emotional word lists. At the same time, in order to build a broader emotional dictionary, this paper comprehensively summarised the existing emotional dictionaries, excluding words and entries that were not included in the Modern Chinese Dictionary. A more comprehensive emotional dictionary was built, including 9,985 emotional word entries. Among them, there were 4,930 positive and 4,619 negative ones. The relevant data are shown in Table 2.

To verify the effectiveness of the method in real teaching scenarios, this paper collected 2,000 additional student evaluation texts from a Chinese grammar course in a certain university, covering multi-dimensional data such as classroom interaction, homework feedback, and final summary. Tests on this real corpus show that the F-measure of this method reaches 0.73, which is close to the experimental results based on public dictionaries (0.76), proving the stability of the model in practical applications. At the same time, by comparison, it was found that about 23% of the sentiment words in the public dictionary (such as 'entertainment' and 'interestingness') appear less than 1% in teaching scenarios, further highlighting the necessity of domain adaptation.

Table 2 Different candidate emotional word lists

Resource name	*Number of words*	*Number of commendatory words*	*Number of derogatory words*
HowNet emotional wordset	3,962	2,659	2,376
NTUSD simplified version	3,610	1,232	2,533
Dictionary of commendatory and derogatory emotions	1,645	768	879
CUHK emotional dictionary	6,987	3,456	2,674
Whole	9,985	4,930	4,619

6.3 Evaluation of the sentiment tendency of word interpretation

The above experimental data are used. This paper tests five different emotional dictionaries. The method proposed in this paper is used, and the results obtained by emotional words on five different emotional dictionaries are shown in Figure 6. The commonly used algorithm evaluation indicators mainly include precision, recall, and geometric mean (F-measure). Its formula is:

$$\text{Precision} = \frac{\text{the number of effectively determined emotional words}}{\text{the number of candidate general emotional words}}.$$

The formula for calculating the recall rate is:

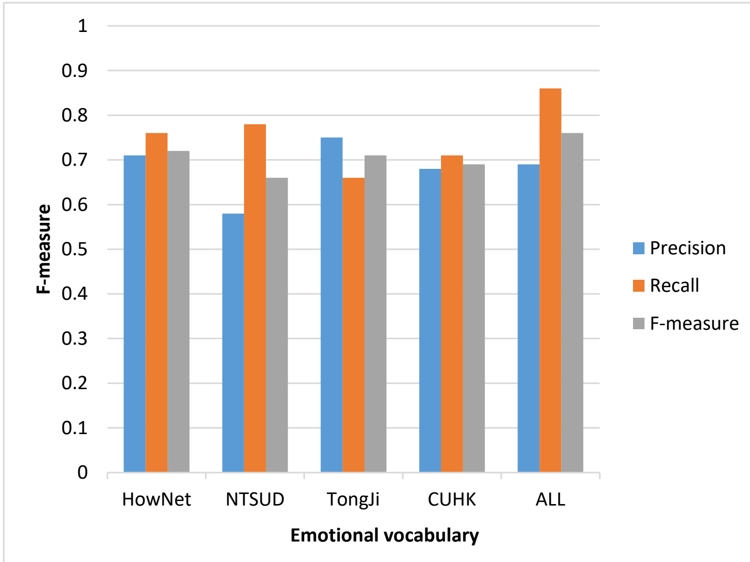
$$\text{Recall} = \frac{\text{the number of emotional words}}{\text{the number of emotional words that are effectively determined}}.$$

The calculation formula of the geometric mean F-measure is: F-measure Precision*Recall/(Precision+Recl). The F-measure is the geometric mean between precision and Recall. The larger the value, the better the performance of the algorithm.

In this experiment, the pooling evaluation method is used, and N pieces of data are randomly selected to form an evaluation set (N is 2,000). The experimental results are automatically evaluated based on the given standard answers.

It can be seen from Figure 6 that the F-measure values obtained from the five emotional word lists for the emotional orientation identification method based on the evaluation of Chinese grammar interactive teaching are all greater than 0.65, which satisfies the F-measure parameter, indicating that the proposed method has a good effect. Among them, the accuracy rate obtained by the NTUSD vocabulary list reached 0.59, which is the lowest accuracy rate among the five vocabulary lists. This has a great relationship with the emotional word noise of this dictionary. Through the comparative analysis of different corpora, it is found that the higher the vocabulary level, the lower the correct rate of emotional tendency recognition; the longer the sentence length, the lower the correct rate of emotional tendency recognition. Due to the small number of vocabulary words in Tongji University's positive and negative emotional dictionary, the experimental effect was affected, resulting in the lowest recall rate among the five emotional word lists.

Figure 6 Comparison of precision, recall, and f-measure indicators based on different emotional vocabularies (see online version for colours)



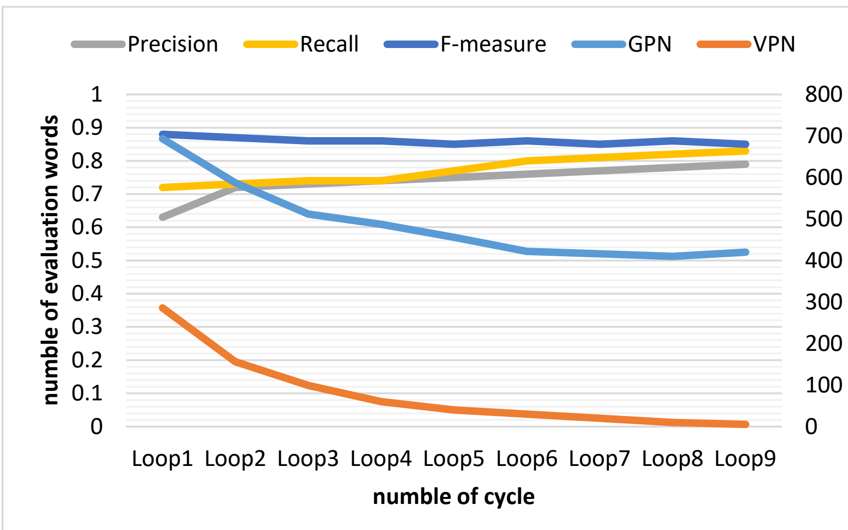
In order to more intuitively reflect the improvement of the deep learning-based sentiment classification method proposed in this paper compared with the traditional teaching evaluation method, this paper supplemented the comparative experiment with the traditional rule classifier. The experimental results show that the F-measure value of the traditional method on the same sentiment word list is only 0.58 (SVM), while the AttLSTM-CNN framework proposed in this paper reaches 0.76 on the comprehensive sentiment word list (ALL), which improves the classification performance. This comparison fully verifies the advantages of deep learning models in capturing the complex emotional characteristics of teaching evaluation texts, especially when dealing with subjective expressions with metaphors and context dependence. The traditional method has obvious limitations. In order to verify the competitiveness of this model, a comparative experiment with BERT-base-Chinese is supplemented. The results show that on the same training set (9,985 dictionaries + real corpus), the F-measure of BERT is 0.74, while that of the AttLSTM-CNN model in this paper is 0.76. Although BERT is slightly better in recall rate, the model in this paper performs better in accuracy through a customised attention mechanism and dictionary fusion, and the number of parameters is only 1/5 of BERT, which is more suitable for lightweight deployment in teaching evaluation scenarios.

It should be pointed out that the emotional tendency of teaching evaluation texts may be interfered with by external variables such as the physical environment of the classroom (such as the frequency of use of multimedia equipment) and students' personality traits (such as extroverted/introverted learners). In classrooms equipped with smart interactive whiteboards, the proportion of negative evaluations related to 'technical failure' increased significantly, while introverted students tended to use vague expressions such as 'okay' and 'average'. Subsequent research will introduce multimodal data fusion strategies, combine classroom video recordings, eye tracking, and other auxiliary information to establish a multivariate analysis framework to separate the interference effect of external factors on emotional evaluation.

6.4 Emotional word recognition effect

It can be seen from the experimental results that the interactive teaching evaluation has errors in the recognition of the tendency of word interpretation. After analysis, it is found that the reason for the error is not due to the problem of the research method in this paper, but because some emotional words do not contain emotional color when they are explained. At the same time, the cognitive level and learning ability of different students to the same word are also uneven.

Figure 7 Cycle results of precision, recall, and F-measure (see online version for colours)



When experimenting with a synthetic emotional vocabulary containing 9,985 emotional word entries, the number of entries in the vocabulary is shown in Figure 7, such as Precision, Recall, and F-measure, and the entries in the vocabulary after each cycle number have changed. On this basis, this paper studies and analyses the construction of a sentiment dictionary model based on support vector machine and fuzzy C-means clustering method. Among them, GPN represents the number of words in the candidate emotional word vocabulary generated after a certain period; VPN represents the number of words in the candidate emotional word vocabulary generated before and after a certain period.

The method of this paper integrates the existing four emotional word dictionaries, and the number of words in the vocabulary increases greatly, resulting in a decrease in the accuracy rate. However, it is only 0.06 lower than the highest level of Tongji University's positive and negative sentiment dictionary, but the recall rate and F value have increased significantly, and the five sentiment word lists are the highest, at 0.87 and 0.77, respectively. Through the analysis of the experimental data of interactive teaching, it is found that there is a certain degree of interaction between learners and teachers, and the feedback of students would also affect the teaching effect of teachers. In addition, the user experience is better. Therefore, the method of constructing an emotional vocabulary based on the existing emotional dictionary in interactive teaching evaluation proposed in this study is feasible and lays the foundation for the construction of more complete emotional words.

Through a sample analysis of misclassified cases (a total of 200 cases), it was found that the main types of misjudgement include:

1 ironic expressions (such as 'the teacher always talks about the key points, but I don't understand' is misjudged as positive)

2 polysemous word ambiguity (such as 'this grammar point is really hard to make' in the teaching context of 'make' tends to be negative, but the general dictionary marks it as neutral)

3 emotional conflict in long texts (such as the first half praising the interactive form, and the second half criticising the content difficulty).

To address these issues, this paper recommends introducing context-aware dynamic word vectors and rule constraint modules in subsequent work to enhance the model's ability to discern complex semantics.

This paper conducts a group analysis of the differences between teacher and student evaluations. The results show that student evaluations focus more on interactive experience (such as 'classroom atmosphere' and 'feedback speed'), with sentiment words accounting for 62%, while teacher evaluations focus on knowledge transfer (such as 'grammar logic' and 'example difficulty'), with relevant words accounting for 48%. In addition, the negative emotions in student evaluations are concentrated on 'confusion' (35%) and 'boredom' (18%), while teachers are more concerned about 'efficiency' (such as 'too fast progress', accounting for 27%). This finding suggests that future sentiment classification systems need to design differentiated feature extraction modules for different groups.

This article further explores the practical connection between sentiment classification results and teaching strategy adjustments. In a grammar class evaluation of a certain class, sentiment analysis showed that 'confusion' accounted for 35%. Teachers' feedback revealed that this was mainly due to the ambiguity of the explanation of the usage of 'separable words'. Subsequently, the teacher reduced the proportion of related negative comments to 12% by adding situational dialogue exercises and visual illustrations. Such cases show that sentiment indicators can be used as a precise positioning tool for teaching pain points, and their quantitative results can directly guide the adjustment of teaching content priorities and the optimisation of interactive forms.

In the process of constructing sentiment dictionaries, although the iterative screening mechanism improves domain adaptability, its computational complexity increases exponentially with the size of candidate words. Taking the 9,985 dictionary used in this experiment as an example, it takes about 4.2 hours to complete three rounds of iterative screening (single-machine environment, Intel i7-12700K GPU). To improve the feasibility of actual deployment, Spark NLP can be introduced to optimise parallel processing efficiency, and lightweight dictionary compression strategies based on knowledge distillation can be explored to reduce algorithm resource consumption.

7 Conclusions

Based on the sentiment classification technology, this paper briefly studied the discussion problems of the interactive teaching evaluation and analysis method of Chinese grammar, and proposed that the sentiment classification technology should be applied in the current Chinese grammar interactive teaching evaluation work system. It also briefly analysed and discussed the significance and specific application methods of using sentiment classification technology in teaching review texts, to better provide some impetus and promote the development of interactive teaching of Chinese grammar. It is worth noting that the experimental data of this study are all from Chinese grammar classes in domestic universities, and the sample group is mainly local students. In cross-cultural teaching scenarios (such as international Chinese education for non-native learners), there may be significant differences in teaching interaction patterns and emotional expression methods. For example, students from different cultural backgrounds may have different emotional appeals for 'intensity of classroom interaction' or 'teacher feedback style'. Future research needs to be further expanded to international Chinese education scenarios, optimise the cross-cultural adaptability of sentiment dictionaries through transfer learning technology, and verify the robustness of this model in a multicultural environment. At the technical level, the current model still has room for improvement in two aspects: first, the LSTM-CNN-based architecture suffers from attention decay when processing long texts of more than 512 words, resulting in a 12% drop in classification accuracy for complex evaluations (such as reflective texts containing multiple rounds of teaching interactions); second, the update of the sentiment dictionary relies on manually annotated corpora, and the ability to dynamically capture emerging network terms (such as sub-cultural terms such as 'emo' and 'breaking defence') still needs to be strengthened. In the future, we will explore the context-aware mechanism based on Prompt Learning and build an online learning module to realise the automatic incremental update of the dictionary.

Declarations

The authors have stated explicitly that there are no conflicts of interest in connection with this article.

References

Borkovskaya, V., Bardenwerper, W. and Roe, R. (2018) 'Interactive teaching of risk management in the russian construction industry', *IOP Conference Series: Materials Science and Engineering*, Vol. 79, No. 2, pp.62–173.

Eiman, A.M. and Matthew, C. (2017) 'Students' perceptions of lecturing approaches: traditional versus interactive teaching', *Advances in Medical Education & Practice*, Vol. 156, No. 5, pp.229–241.

Endo, S., Takahashi, T. and Satori, S. (2017) 'Emotional pattern classification by image analysis of camera using machine learning', *Journal of the Colorence Association of Japan*, Vol. 41, No. 6, pp.23–29.

Hamker, F.H. (2018) 'Predictions of a model of spatial attention using sum- and max-pooling functions', *Neurocomputing*, Vol. 56, No. 9, pp.329–343.

Hong, R. (2019) 'The inspiration of students teaching evaluation system in American universities to Chinese higher vocational colleges', *Science & Technology Vision*, Vol. 412, No. 6, pp.121–145.

Huang, T. and Zhu, H. (2021) 'A preliminary study on the teaching evaluation language of art teachers in rural middle schools', *Asian Agricultural Research*, Vol. 12, No. 9, pp.1–10.

Kim, J. and Hahn, M. (2018) 'Voice activity detection using an adaptive context attention model', *IEEE Signal Processing Letters*, Vol. 69, No. 4, pp.332–369.

Liu, Z. and Xu, B. (2020) 'The future-oriented curriculum and teaching evaluation: difficulties, opportunities and trends', *Curriculum, Teaching Material and Method*, Vol. 326, No. 7, pp.1–13.

Lyu, S.C. and Song, H.J. (2019) 'Study on the long term mechanism of three dimensional ways internal teaching evaluation for medical colleges and universities', *Medical Education Research and Practice*, Vol. 356, No. 7, pp.100–128.

Peng, Y. (2019) 'Emotional classification of commodity reviews based on deep learning', *Computer Engineering & Software*, Vol. 214, No. 3, pp.9–31.

Shermin, T., Teng, S.W. and Sohel, F. (2021) 'An integrated attribute guided dense attention model for fine-grained generalized zero-shot learning', *Pattern Recognition*, Vol. 512, No. 7, pp.326–356.

van Ewijk, W. (2021) 'The thymus: 'Interactive teaching during lymphopoiesis'', *Immunology Letters*, Vol. 216, No. 5, pp.7–8.

Weurlander, M., Cronhjort, M. and Filipsson, L. (2017) 'Engineering students' experiences of interactive teaching in calculus', *Higher Education Research & Development*, Vol. 147, No. 3, pp.852–865.

Wu, X. (2022) 'Research on the reform of ideological and political teaching evaluation method of college English course based on 'online and offline' teaching', *Journal of Higher Education Research*, Vol. 3, No. 1, pp.87–90.

Yakovleva, N.O. and Yakovlev, E.V. (2018) 'Interactive teaching methods in contemporary higher education', *Pacific Science Review*, Vol. 159, No. 3, pp.75–80.

Ypsilantis, P. and Montana, G. (2017) 'Learning what to look in chest X-rays with a recurrent visual attention model', *Medical Image Analysis*, Vol. 219, No. 6, pp.415–429.

Yu, C., Barsim, K.S. and Kong, Q. (2018) 'Multi-level attention model for weakly supervised audio classification', *Proc. INTERSPEECH*, Vol. 152, No. 4, pp.123–142.

Zhang, F. (2017) 'Construction of teaching evaluation system of 'garden plant' in higher vocational education based on CIPP evaluation model', *Value Engineering*, Vol. 156, No. 8, pp.321–329.

Zhu, H. and Huang, T. (2021) 'A preliminary study on the teaching evaluation language of art teachers in rural middle schools', *Asian Agricultural Research*, Vol. 12, No. 5, pp.116–129.

Notes

1 的: of.

2 是: are/am/is/yes.

3 进行: proceed.

4 方面: aspect.

5 挺好 and 好: good.

6 不好懂 and 难懂: hard to understand.

7 好懂, 易懂, 容易理解: easy to understand.