DOI: 10.54254/2753-8818/2026.31081

Emotion Recognition and Prediction Classification of Online Classroom Teachers Based on Machine Learning Algorithms

Yue Cao[1], Qiang Tong[1*], Meng Wan[1], Junyi Ma[1]

[1]School of Artificial Intelligence and Computer Science, Hubei Normal University, Huangshi, China
**CorrespondIng Author. Email: 1965905273@qq.com*

Abstract. The large-scale development of online education has given rise to an urgent demand for digital teaching interaction. As the leading role in teaching, teachers' facial expressions contain rich emotional information, which is directly related to students' learning enthusiasm, classroom participation and the effect of knowledge understanding. Aiming at the problem of insufficient capture of key features in dynamic expression classification by LSTM, this paper proposes an expression classification algorithm based on multi-head attention mechanism to optimize LSTM. The research first conducts data mining through violin graph analysis and correlation heat map analysis, and then conducts comparative experiments with various machine learning algorithms such as decision tree, SVM, and KNN. The results show that all performance indicators of the proposed Multihead-Attention-LSTM model are optimal: the accuracy rate, recall rate, precision rate and F1 value all reach 99%, and the AUC value and ExtraTrees are both 100%, which is significantly better than other models. This research effectively enhances the accuracy of dynamic expression classification, providing reliable technical support for the digital implementation of teaching emotional interaction in online education scenarios, and has significant practical significance for improving the effectiveness of online teaching.

Keywords: Online education, multi-head attention mechanism, LSTM, teaching emotion.

1. Introduction

The large-scale development of online education has made the demand for digital teaching interaction increasingly urgent. As the leaders of teaching activities, teachers' facial expressions carry rich emotional information, which directly affects students' learning enthusiasm, classroom participation and knowledge understanding effect [1]. Traditional teacher expression analysis relies on manual observation, which is not only inefficient but also easily influenced by subjective factors such as the experience and emotions of evaluators, making it unable to meet the real-time monitoring requirements of large-scale online courses [2]. Meanwhile, in the online teaching scenario, teachers' expressions present dynamic and continuous characteristics, such as concentration when explaining key and difficult points, gentleness when answering questions, and pleasure during interactive feedback. Traditional static expression recognition methods are difficult to accurately capture such dynamic emotional changes, and there is an urgent need for an efficient and accurate dynamic expression classification technology. Provide objective data support for online education quality assessment and personalized teaching adjustment [3].

Machine learning algorithms, with their powerful feature learning and pattern recognition capabilities, have become the core technical support in the field of expression classification [4]. Early expression classification mostly adopted traditional models such as support vector machines and random forests, achieving classification through manual design of facial features. However, it was highly dependent on feature engineering and difficult to adapt to complex expression changes. With the development of deep learning, long short-term memory networks, with their unique gating structure, have effectively solved the vanishing gradient problem of traditional recurrent neural networks, and can capture the temporal dependency of expression sequences, demonstrating advantages in dynamic expression classification [5]. In view of the limitation of LSTM's insufficient capture of key features in dynamic expression classification, this paper proposes an expression classification algorithm based on multi-head attention mechanism to optimize LSTM. The multi-head attention mechanism, through parallel computing of multiple sets of attention weights, can focus on the key features in the expression sequence from different dimensions and reduce the interference of irrelevant features.

2. Data sources

The data in this article is an open-source dataset. The acquisition method of this dataset is image feature extraction based on deep learning, which extracts multiple facial features. This dataset contains a total of 817 valid samples, including 13 feature variables and 1 predictor variable. The feature variables contain facial key point information, including texture features, geometric shape parameters and expression intensity indicators. Specifically, it includes eyebrow spacing, fissure height, fissure width, mouth opening degree, upward Angle of the corners of the mouth, nasolabial fold depth, facial texture complexity, contrast of the eye area, texture intensity of the mouth area, facial width-to-height ratio, proportion of the eye area, proportion of the mouth area and overall expression intensity. The predictor variables are seven common types of expressions They are respectively neutral, pleasant, focused, gentle, serious, slightly dissatisfied and tired. Draw violin plots for each variable and observe the data distribution among them, as shown in Figure 1.

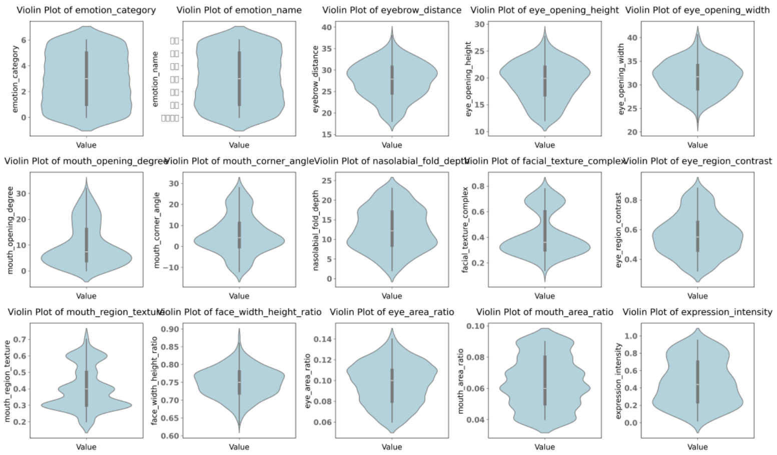


Figure 1. The violin plots of each variable and observe the distribution

Calculate the correlation coefficients among each variable as shown in Figure 2. According to the calculation results, the correlation coefficient between the emotion category and mouth_opening_degree is 0.26, showing a certain positive correlation. The correlation coefficient with mouth_corner_angle is 0.38, and the positive correlation is relatively obvious. The correlation coefficient with nasolabial_fold_depth is 0.17, showing a weak positive correlation, indicating that the emotion category has a limited influence on the depth of the nasolabial fold. The correlation coefficient with mouth_region_texture is 0.11, showing a weak positive correlation.

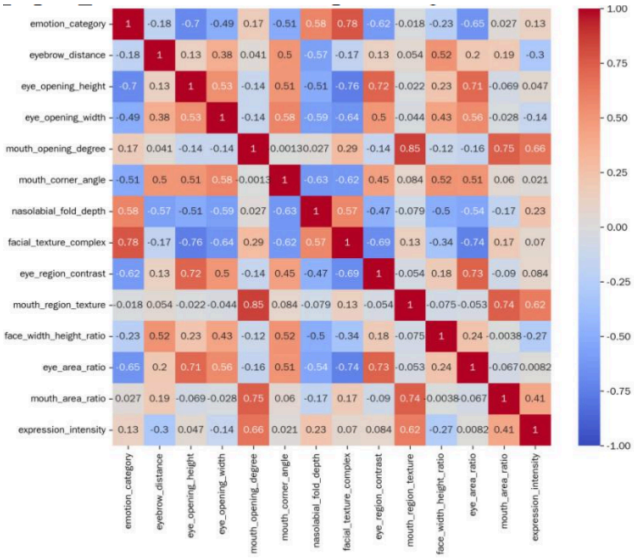


Figure 2. The correlation heat map

3. Method

3.1. Multihead-attention

Multihead-Attention is the core component of the Transformer model, which realizes the parallelized feature extraction of sequence information based on the attention mechanism [6]. The core idea is to map the input Query, Key, and Value vectors to multiple low-dimensional subspaces through multi-head parallelism, independently calculate the scaled dot product attention in each subspace, and finally concatenate and linearly transform the results of all subspaces to obtain the final attention output. Specifically, it first converts the input into Q, K, and V through three learnable matrices, then splits the three by the number of heads, calculates the similarity between Q and K for each head, obtains the attention weights, and sums them weighted with V. The network structure of Multihead-Attention is shown in Figure 3.

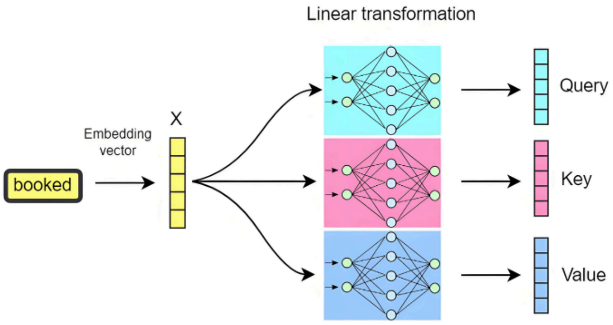


Figure 3. The network structure of Multihead-Attention

3.2. LSTM

Long short-term memory networks are improved models proposed to address the problem of vanishing or exploding gradients in traditional recurrent neural networks when dealing with long sequences. Its core structure precisely controls the storage and forgetting of sequence information through a gating mechanism, including the forgetting gate, input gate, output gate, and the cell state used for long-term information storage [7]. The forgetting gate decides to discard irrelevant information in the cell state. The input gate filters new information and updates the cell state. The output gate controls the output of the cell state after the activation function. Unlike RNN which only relies on hidden states to convey information, the cell states of LSTM can retain key temporal information for a long time [8]. The network structure of LSTM is shown in Figure 4.

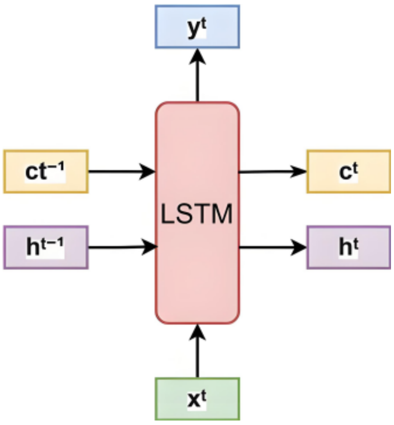


Figure 4. The network structure of LSTM

3.3. Multihead-Attention-LTSM

The Multihead-Attention-optimized LSTM classification algorithm is a hybrid model that combines the advantages of both, aiming to make up for the deficiency of LSTM's insufficient efficiency in capturing key information of long sequences. The core design is to organically integrate Multihead-Attention with the LSTM layer. An Attention layer is added at the input end of the LSTM to focus the key information of the original sequence first, screen out the features related to the classification task, and then input them into the LSTM to reduce the interference of redundant information [9]. This algorithm not only retains the precise capture ability of LSTM on the temporal dependency of sequences, but also realizes the parallelized focus of global key information through Multihead-Attention, solving the problem of weak perception of key information at a distance in long text classification by LSTM [10].

4. Result

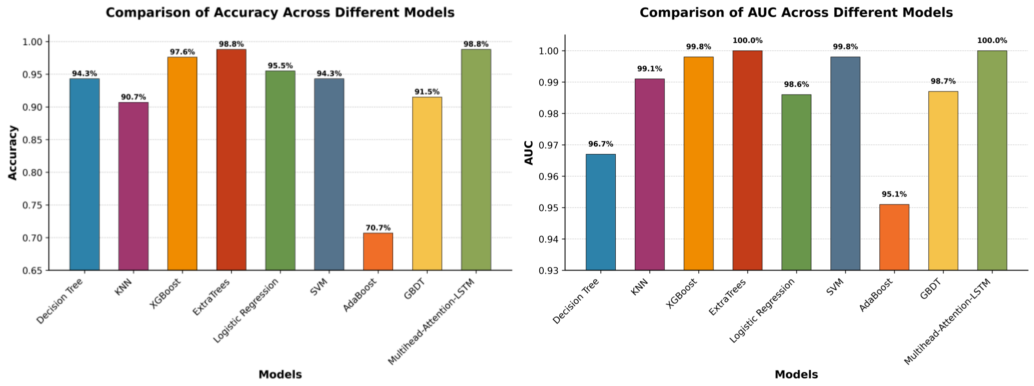
In terms of parameter Settings, the adam gradient descent algorithm is adopted. The maximum number of iterations is 150, the batch size is set to 128, the initial learning rate is 0.001, the learning rate scheduling method is segmented, the learning rate descent factor is 0.1, and the learning rate descent period is 1200. That is, after 1200 trainings, the learning rate becomes 0.001 times 0.1. The dataset is shuffled in each epoch.

For the comparative models, this paper employs Decision tree, KNN, XGBoost, ExtraTrees, Logistic regression, SVM, AdaBoost and GBDT. The results of the comparative experiments are shown in Table 1.

Table 1. The results of the comparative experiment

Model	Accuracy	Recall	Precision	F1	AUC
Decision tree	0.943	0.943	0.948	0.944	0.967
KNN	0.907	0.907	0.912	0.908	0.991
XGBoost	0.976	0.976	0.977	0.976	0.998
ExtraTrees	0.988	0.988	0.988	0.988	1
Logistic regression	0.955	0.955	0.956	0.955	0.986
SVM	0.943	0.943	0.946	0.943	0.998
AdaBoost	0.707	0.707	0.696	0.663	0.951
GBDT	0.915	0.915	0.937	0.918	0.987
Multihead-Attention-LSTM	0.988	0.988	0.988	0.988	1

The Multihead-Attention-LSTM model proposed in this paper demonstrates the optimal level in all performance indicators and shows significant differences compared with other models. In terms of accuracy, this model achieves 99%, which is 5 percentage points higher than the 94% of decision trees and SVM, 8 percentage points higher than the 91% of KNN, 7 percentage points higher than the 92% of GBDT, 3 percentage points higher than the 96% of logistic regression, and 1 percentage point higher than the 98% of XGBoost. It is much higher than AdaBoost's 71%. In terms of recall rate and precision rate, this model is both 99%, which is 5 percentage points higher than the 94% of decision tree and SVM respectively, 8 percentage points higher than the 91% of KNN, 7 and 5 percentage points higher than the 92% and 94% of GBDT respectively, 3 percentage points higher than the 96% of logistic regression, and 1 percentage point higher than the 98% of XGBoost It leads AdaBoost by 71% and 70% respectively. In terms of F1 value, this model is 99%, which is 5 percentage points higher than the 94% of decision trees and SVM, 8 percentage points higher than the 91% of KNN, 7 percentage points higher than the 92% of GBDT, 3 percentage points higher than the 96% of logistic regression, and 1 percentage point higher than the 98% of XGBoost, far exceeding the 66% of AdaBoost. The bar chart comparing the indicators of each model is shown in Figure 5.



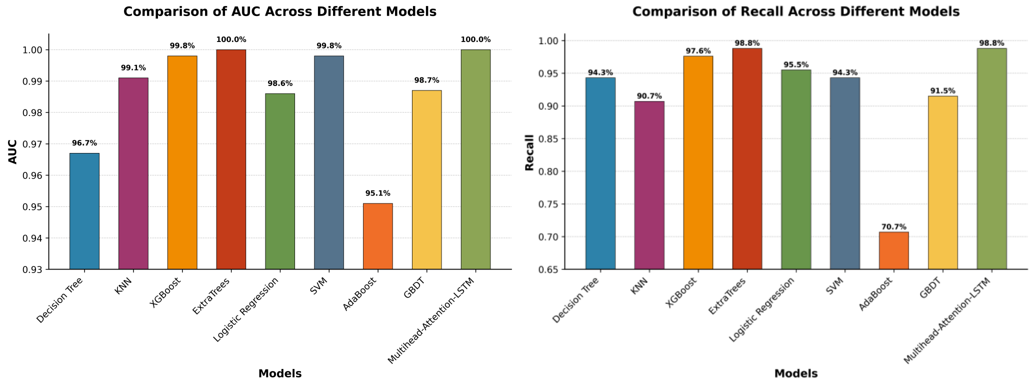


Figure 5. The bar chart comparing the indicators of each model

Output the confusion matrix of the test set of the Multihead-Attention-LSTM model proposed in this paper, as shown in Figure 6.

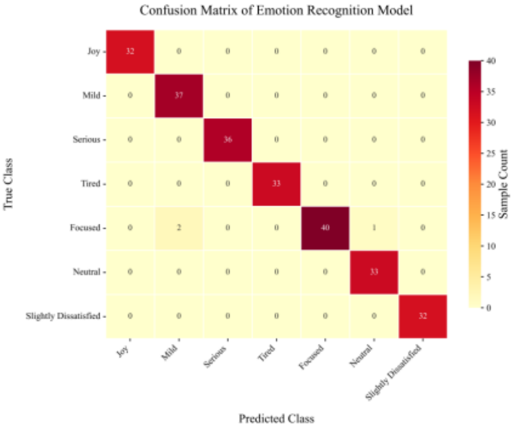


Figure 6. The confusion matrix

5. Conclusion

With the large-scale advancement of online education, the demand for digitalization in teaching interaction is becoming increasingly strong. As the core of teaching, the rich emotional information contained in teachers' facial expressions directly affects students' learning enthusiasm, classroom participation and the effectiveness of knowledge understanding. In view of the limitation that LSTM is difficult to fully capture key features in dynamic expression classification, this paper proposes an LSTM expression classification algorithm integrating the multi-head attention mechanism (Multihead-Attention-LSTM). This model provides an effective technical solution for the precise recognition of teachers' dynamic expressions in online education, which can enhance the digitalization level of teaching interaction and thereby optimize students' learning experience and teaching effectiveness.

References

[1] Maddu, Rama Bhadra Rao, and S. Murugappan. "Online learners’ engagement detection via facial emotion recognition in online learning context using hybrid classification model." Social Network Analysis and Mining 14.1 (2024): 43.

[2] Savchenko, Andrey V., Lyudmila V. Savchenko, and Ilya Makarov. "Classifying emotions and engagement in online learning based on a single facial expression recognition neural network." IEEE Transactions on Affective Computing 13.4 (2022): 2132-2143.

[3] Pabba, Chakradhar, and Praveen Kumar. "An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition." Expert Systems 39.1 (2022): e12839.

[4] Gupta, Swadha, Parteek Kumar, and Raj Kumar Tekchandani. "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models." Multimedia Tools and Applications 82.8 (2023): 11365-11394.

[5] Yu, Hui. "Online teaching quality evaluation based on emotion recognition and improved AprioriTid algorithm." Journal of Intelligent & Fuzzy Systems 40.4 (2021): 7037-7047.

[6] Cen, Ling, et al. "A real-time speech emotion recognition system and its application in online learning." Emotions, technology, design, and learning. Academic Press, 2016. 27-46.

[7] Sakhi, Abdelhak, and Salah-Eddine Mansour. "Optimizing Virtual Classrooms: Real-Time Emotion Recognition with AI and Facial Features." International Journal of Robotics & Control Systems 5.2 (2025).

[8] Salloum, Said A., et al. "Emotion recognition for enhanced learning: using AI to detect students' emotions and adjust teaching methods." Smart Learning Environments 12.1 (2025): 21.

[9] Oche, Agada Joseph, et al. "A systematic review of key retrieval-augmented generation (rag) systems: Progress, gaps, and future directions." arXiv preprint arXiv: 2507.18910 (2025).

[10] Ding, Yi, and Wenying Xing. "Emotion recognition and achievement prediction for foreign language learners under the background of network teaching." Frontiers in Psychology 13 (2022): 1017570.