ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

Research on classroom behavior analysis and optimization strategies based on convolutional neural network in smart learning environment

Jin Lu[1,*]

[1]*Guangdong Key Laboratory of Big Data Intelligence for Vocational Education, Shenzhen Polytechnic, Shenzhen, Guangdong, 518055, China*
[*]*Corresponding author: lujin0808@szpt.edu.cn*

Abstract: *Realizing the analysis and prediction of attention and learning emotion recognition in teacher-student adaptive interaction in a smart teaching environment is a major development need in the fields of intelligent guidance, teacher classroom evaluation, and classroom stress analysis. However, classroom education and online education big data are characterized by cross-modality, multi-source heterogeneity, content redundancy and structural confusion, and the secrecy of teacher-student relationship between different types of courses is high, which brings great challenges to the analysis and prediction of attention and learning emotion recognition in teacher-student adaptive interaction. This paper is based on the research idea of "sentiment analysis-data modeling-group discovery-behavior prediction", and innovatively uses the core algorithm of convolutional neural network to overcome the key problems of attention and learning emotion recognition and prediction in teacher-student adaptive interaction in the intelligent teaching environment. This paper presents an innovative approach to the analysis and prediction of attention and learning emotions in teacher-student adaptive interaction in a smart teaching environment. The method proposed in this paper achieves 96.88% accuracy in practical application at the expense of computation time. This study is expected to form a key method for the analysis of teacher-learner adaptive interaction states driven by multimodal data in a smart teaching environment, which provides optimized strategies and theoretical support for personalized smart teaching, and has high theoretical significance and application value.*

Keywords: *Classroom behavior analysis, Convolutional neural network, Smart learning environment*

1. Introduction

Classroom teaching is the main channel for students' development. Using intelligent technology tools to observe and analyze classroom teaching behavior can optimize teachers' classroom teaching behavior, promote teachers' professional development, and enhance teaching efficiency and quality. Meanwhile, teaching quality is the core of smart education, and how to improve the quality and efficiency of teachers' and students' teaching activities in classroom teaching contexts is an important element of educational technology. In the boom of ChatGPT/GPT-4 and other large-scale AI sweeping the world, people have deeply recognized the great leap and breakthrough development that AI, as a transformative technology and key force in economic and social development, will bring to the global industry, profoundly affecting the future world competition pattern and the development of smart education.

In the process of education and teaching, the attention and learning emotions of teachers and students in adaptive interaction are important factors affecting the efficiency of classroom learning, and are the core indicators for evaluating teaching quality. In the era of rapid development of intelligent technology and deep promotion of digital transformation in all walks of life, it is technically and theoretically feasible to solve the problem of difficult identification of dynamic data of teacher-student interaction in the process of intelligent teaching through digital means and carry out intelligent monitoring and assistance of classroom teaching. From the application level, with the integration of artificial intelligence, meta-universe, cloud computing, big data and other intelligent technologies into the whole process of teaching, the digital transformation of education technology has been fully launched, and smart education has gradually emerged, bringing more efficient, intelligent and personalized teaching experience. Smart teaching usually consists of a series of complex teaching

ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

activities and links, and the teacher-student interaction information generated in each activity has the characteristic of being closely integrated with the context. Analyzing the attention and learning emotion data related to learning and building a collection model based on multimodal data can ensure accurate perception of changes in classroom teaching quality. In the classroom learning process accompanied by a series of behavioral activities and psychological activities, meaningful learning is needed to build on the basis of knowing oneself, teachers or educational agents should care about their students' progress process and mental path changes, which can help develop students' sense and can encourage students to care about their own progress. In order to fully understand how the learner's learning state evolves during the learning process, it is important to take advantage of emotion recognition technology to develop intelligent emotion systems from the perspective of students, teachers, and smart education researchers to identify and respond appropriately to their emotional changes. At the academic level, the easy access to teacher-student adaptive interaction data supported by smart technologies provides an effective way to track and capture changes in noticing smart teaching activities. However, limited to the characteristics of multimodal, heterogeneous, incomplete, and strongly correlated dynamically generated data in the process of smart teaching, it is difficult to effectively collect and deeply analyze them by traditional methods. Establishing a unified data collection specification and analysis mechanism to efficiently collect and scientifically analyze process data has become an urgent problem to be solved in current smart education research. Therefore, it has become urgent to explore new theories, technologies, methods and approaches to realize classroom teaching quality improvement while continuing to study and improve key technologies of education informatization in depth.

2. Literature review

Multimodal identification of teacher-student adaptive interactions is the key to modeling data of smart teaching contexts. With the support of smart technologies, easy access to teacher-student adaptive interaction data provides an effective way to track and grasp the changes in teaching activities[1]. With the widespread use of sensor technologies and the integration with smart wearable devices, a number of new technologies and tools, each with its own characteristics, such as CNN, have been designed and developed to provide great convenience for data processing and analysis.

The current state of research on CNN detection techniques. CNNs have been used in a wide variety of fields and tasks. Online social networks such as Facebook, Twitter and WeChat reveal similar interests among online users, and CNNs based on online social behavior can effectively infer relationships between users and user preferences, and are used for tasks such as spammer detection and crisis response [2]. Neuroscience is the study of the nervous system and the brain, and with recent developments in brain mapping and neuroimaging techniques, the brain is also beginning to be modeled as a network, and CNNs based on brain networks can help identify functional parts of the brain that play a role or have pathologie[3]. CNN-based image understanding generates better semantic descriptions of images by introducing educational aids for teaching and learning[4]. Recommendation is usually based on the information in the user's purchase or browsing history to build a profile of the user's interests and then recommend similar items to the user to solve the user information overload problem, and association discovery by introducing the concept of association generates high-quality recommendation results by effectively detecting the relationships between nodes[5]. Link prediction deals with missing connections and predicts possible future connections by analyzing the observed network structure and external information, and link prediction by introducing the association concept analyzes the probability of predicting links between nodes by designing association-specific similarity matrices[6].

Current status of research on attention recognition theory in classroom teaching environment. Objective and accurate assessment of students' learning states such as concentration, emotion and relaxation in real classroom situations is related to education quality monitoring and classroom teaching reform. Traditional methods of teacher-student interaction attention recognition in classroom teaching environment include direct observation method[7], questionnaire test method[8], homework test method[9] and computer-assisted test method[10]. With the help of wearable devices, portable sensors, web-based media tools and various information software, learning data from interactions can be captured and recorded flexibly. However, with the popularity and application of computer vision, machine learning, and brain interfaces, video analysis, ECG data analysis, EEG data analysis, and eye movement analysis are used as recognition methods based on sensing technologies.

The current state of research on learning emotion and attention perception and quantification techniques: the identification of affective states occurs in a series of interactions with learning

ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

technologies, covering all teaching and learning venues, both online and offline. Chengquan et al. argue that the cognitive domain mainly includes the application of four aspects of learners' attention, classroom engagement, cognitive load and creativity[11]. Meanwhile, some other scholars consider the affective domain as the identification and application of some typical and common learning-related emotions, including confidence, hesitation, happiness, anxiety, excitement, calmness, engagement, confusion, boredom, and hope. Since the boom of deep learning techniques, the main application of sensing data-based learning analytics in motor skills is human action recognition, which identifies learners' body movements and sequences by two types of methods based on visual sensing and inertial sensing to provide understanding of learning states[12].

Current status of research on multimodal-based educational teaching aids. There are four methods of learning emotion recognition: emotion analysis based on psychological scales, emotion analysis based on physiological signals, emotion analysis based on speech and expressions, and emotion analysis based on text data. The four emotion recognition methods differ in terms of data source, naturalness, authenticity and engineering volume. Among them, measuring the change of learners' attention by eye changes is a more common measurement method. The learner's gaze information is tracked using eye movement to obtain data on gaze, sweep, blink, and pupil diameter, which can be used to distinguish the learner's state of attentional focus.The effect of task execution time on attention and baseline pupil diameter was studied in detail by Van et al[13]. To estimate user engagement and engagement state from eye-movement information, an empirical study by Li et al. demonstrated that the duration of gazing at a target object can be used to measure user engagement and designed four engagement estimation methods[14].Grafsgaard et al. used features such as facial eyebrows to obtain learners' learning frustration and thus determine learners' cognitive state. Liu et al. correlated eyebrow, eye and mouth features to correspond to three attentional states, demonstrating a correlation between facial information and concentration[15].

This study proposes to carry out research on the identification of attention and learning emotions in teacher-student adaptive interaction based on multimodal data in the classroom environment, working on the following four aspects. First, we will study the perception and quantification mechanism of classroom learning attention, use multimodal data recognition as the main means to achieve real-time concomitant acquisition of attentional behavior and emotion analysis, and explore the elemental characteristics that effectively reflect the level of attention. Secondly, by exploring the pre-processing method of multimodal data of classroom behaviors based on multi-view video data in classroom teaching context, we build a dynamic generative data acquisition and analysis model of intelligent teaching and establish a large-scale database of teaching behaviors and learning emotions. Thirdly, we conduct research on collaborative modeling methods based on association detection, design semantic feature recognition models for multimodal primitive data such as speech and image based on saliency and depth models and consistent representation models for multimodal data. Finally, we explore the construction and application research of an assisted classroom teaching service system based on multimodal data recognition, develop an attention and learning emotion feedback and intervention system based on multimodal data in real classroom situations, and provide optimal strategies and adaptive assistance to promote self-regulated behaviors in classroom learning.

3. Method

3.1 An improved convolutional neural network

Based on multi-viewpoint video data in classroom teaching context and relying on CNN-related algorithms, this paper focuses on the enhanced technology of teacher-student teaching collaborative analysis for classroom teaching, mainly including teaching object somatosensory part detection and recognition technology, teaching object somatosensory-emotion collaborative understanding, multi-viewpoint classroom teaching object target matching, and adaptive assisted education service technology. By detecting and emotionally analyzing the multimodal data of teaching objects, we construct an assisted classroom teaching service system and application based on multimodal collaborative computing, and promote the development of teaching mode in digital intelligence environment. In this paper, we propose to explore the optimal strategy of semantic modeling of teaching effect through multimodal collaborative data collection, modeling and correlation analysis, and analyze the actual situation of teaching quality evaluation in China, and develop an adaptive classroom assisted teaching system based on multimodal data collaboration, and conduct functional experimental tests through real teaching platform. Based on the research idea of "emotion analysis-data

ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

modeling-group discovery-behavior prediction", we conducted an in-depth research on the identification and analysis of attention and learning emotion in teacher-student adaptive interaction under the intelligent teaching environment. The research structure is shown in Figure 1.

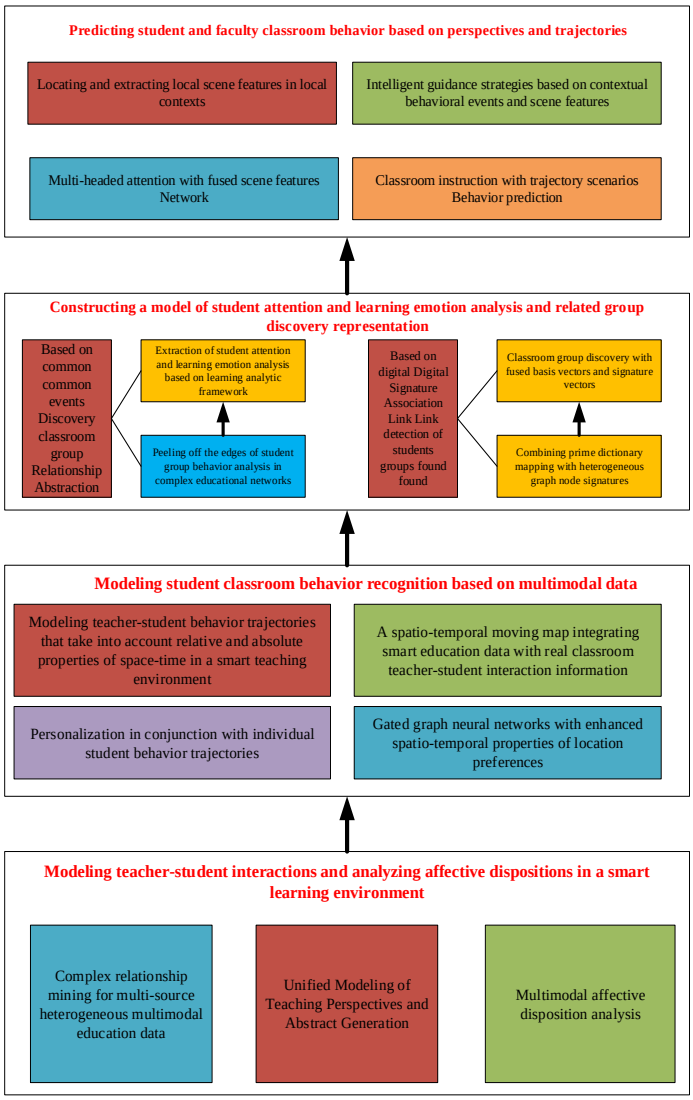

Figure 1: Proposed method

Sentiment analysis. Combining multi-source heterogeneous classroom teaching big data and online teaching big data, taking into full consideration their heterogeneous characteristics, realizing the structured representation of teacher-student behavioral information, and giving solutions for complex correlation relationship mining, unified modeling and summary generation of viewpoints, and multimodal sentiment tendency analysis of cross-teaching mode behavioral data.

Data modeling. In view of the real situation that online learning behavior information is not sufficiently integrated with real classroom teaching, we study the accurate modeling methods for attention and learning emotion identification under the teacher-student adaptive interaction mode.

Group discovery. Combining classroom teaching multimodal learning behavior information data, establishing a classroom character relationship network with teacher-student teaching as the starting point, constructing a classroom behavior relationship network based on group discovery of common events, and exploring a local CNN algorithm with teacher-student as the core combined with digital signature.

Behavior prediction. Combine the viewpoint summary, tendency analysis and accurate modeling of classroom behavior trajectory of teacher-student adaptive interaction behavior events for classroom teaching, and also combine the contextual scene features of teacher-student interaction for events to construct the constraint relationship inherent in classroom behavior events and corresponding trajectory

ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

scene features, so that they can correct the prediction model and improve students' attention and learning in classroom environment Emotion recognition effect.

3.2 Model Architecture

3.2.1 Learning analysis framework based on multimodal data

To study the perception and quantification mechanism of classroom learning attention, to realize real-time concomitant acquisition and affective analysis of attentional behavior with multimodal data recognition as the main means, and to mine the elemental features that effectively reflect the attention level. With the emergence of new technologies such as big data and Internet+, classroom teaching videos also present big data and multi-source heterogeneity. Starting from the recognition theory of attention and learning emotion, this paper systematically studies traditional attention recognition methods such as direct observation method, questionnaire test method, homework test method and computer-aided test method, as well as attention recognition methods, processes and principles based on auxiliary sensing technologies such as video analysis method, ECG data analysis method, EEG data analysis method and eye movement data analysis method, to build a learning analysis based on multimodal data We will build a framework for learning analysis based on multimodal data, and realize teacher-student interaction modeling and emotional disposition analysis in a smart learning environment. The specific architecture is shown in Figure 2.

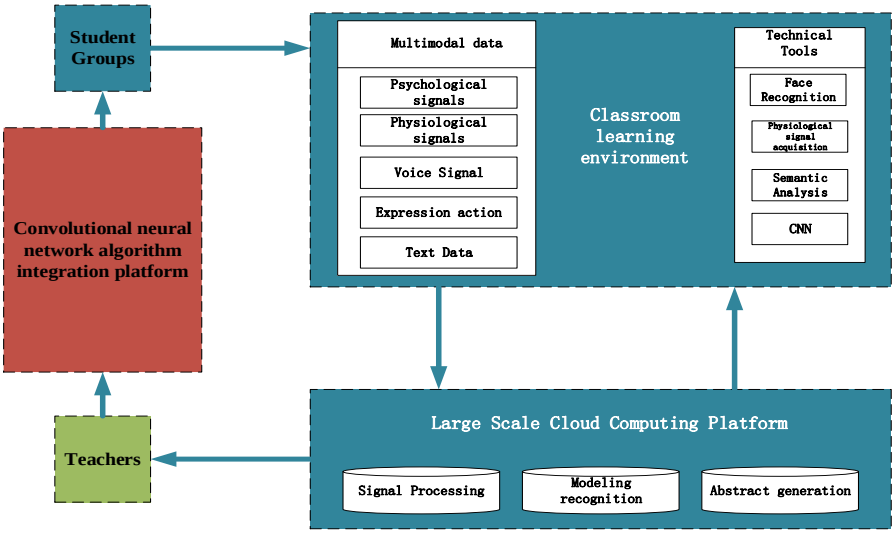

Figure 2: Learning analysis framework based on multimodal data

We collected and organized relevant literature, conducted literature research and analysis, systematically sorted out the main classroom teaching behavior analysis methods at home and abroad, and conducted comparative research on classroom teaching behavior analysis methods.

Through comparative analysis of typical classroom teaching behavior coding systems, we comprehensively examine the differences of teaching behaviors in the digital classroom environment and construct a coding system for teaching behaviors in the digital classroom environment; based on the performance of teachers' and students' teaching behaviors in the digital classroom context, we conduct systematic research on teachers' and students' explicit behaviors and nonverbal behaviors represented by gestures with multiple representations, explore the We explore the classification criteria of teaching behaviors in the digital classroom environment, build a classroom teaching behavior coding system with four categories and 16 behaviors, and carry out complex relationship mining of multi-source heterogeneous multimodal educational data.

Alternatively, the model is trained using the face attribute dataset and the face expression dataset, using k batches of face attribute data followed by 1 batch of face expression data. k is equivalent to the objective function in the relatively small case:

$$l = -\sum_{i}^{C} t_i \log(s_i) - \sum_{k}^{E} t_k \log(f(o)_k) \tag{1}$$

ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

$$f(o)_k = \frac{e^{o_k}}{\sum_k^E e^{o_k}}$$

Here: (2)

Where t_i is the label of the input face in b face attributes, s_i is the output of the face attributes of the model, and t_k is the label of the input data in face expression classification, the specific experimental results are shown below.

This paper proposes a multimodal data summary generation method combining the dual ideas of machine translation and text alignment, which systematically combines the capability of summarizing texts from high-quality language resources in classroom teaching and the advantage of text alignment with circular consistency constraints into a unified structured summary generation model, thus not only effectively improving the recognition of common viewpoints in the misleading and redundant multimodal heterogeneous data, but also supporting quality of classroom summary generation in the context of low-quality resource languages.

In this paper, we propose a novel framework for classroom learning analysis, in which initial answer information is used for visual question-answer analysis, while answers in the heavy-attention mode are used to help optimize the learning of visual attention weights, and answers with initial simple responses are used to compute attention weights for question images. Meanwhile, this paper defines an attentional consistency loss model to help model the key visual attention weights of learning images with question images by minimizing the attentional consistency loss in terms of the distance between visual attention mapping maps learned in a smart teaching environment for multimodal affective disposition analysis and accurate attentional map mapping.

3.2.2 A dynamic generative data collection model for context-aware smart teaching

Establish dynamic generative data collection model, description specification, storage method and exchange mechanism of wisdom teaching based on context perception, and build a large multimodal model. Teaching context acquisition is the key to contextual data modeling. The teacher-student interaction data of teaching process mainly consists of multi-source heterogeneous cross-media data, which are multi-angle descriptions of the same thing, and they are closely related and tend to have the same potential meaning, but because they are often multi-source heterogeneous, it is difficult to model and analyze them uniformly with the same model. In this paper, we dig deeper into the complex correlations of teaching contexts, decompose the specific activities of smart teaching into different stages, innovatively divide the smart teaching contexts of dynamic generative data collection into six categories, such as user, task, location, time, equipment, infrastructure, etc., and decompose them into specific behaviors in combination with each stage, taking the specific behaviors of teachers or students or the situations related to specific behaviors as trigger moments. Obtain the generative context of dynamic generative data of smart teaching and realize the representation of character trajectory based on spatio-temporal movement map.

The softmax loss function is analyzed. softmax separates inter-class features by maximizing the posterior probability corresponding to the correct label. The formula is:

$$L = \frac{1}{N}\sum_{i=1}^{N} -\log p_i = \frac{1}{N}\sum_{i=1}^{1} -\log \frac{e^{f_{yi}}}{\sum_{j=i}^{C} e^{f_i}} \tag{3}$$

Where p_i represents the corresponding posterior probability, Nis the total number of samples trained, C is the total number of classifications, and $f_i$denotes the output of the fully connected layer:

$$f_i = W_i^T x + B_i \tag{4}$$

Teaching behavior recognition algorithm based on object detection, including student hand raising, standing, sleeping, yawning, student attention, student positioning, etc. Facial expression recognition algorithm based on motion unit AU detection and tensor sparse representation. The algorithm incorporates temporal information of motion units, association information between motion units, more accurate face detection and face alignment methods, and feature extraction based on tensor sparse representation. Starting from AU, the facial expressions of teachers and students are described by different combinations of AU. Based on the distribution differences between different AU and the

ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

temporal information occurring for the facial expression action unit, the temporal features are incorporated into the recognition network, and the recognition of facial expressions of teachers and students is also performed based on the correlation between AU and expressions. The PPG signal data of the subject is measured from the wrist by the sensing device, which can effectively suppress the influence of ambient light on the signal and provide data such as basic PPG waveform and heart rate detection. Meanwhile, by analyzing the student's heart rate variability and obtaining the vagal tone of the heart, data on the student's sustained attention change can be obtained. Based on the multimodal data, taking into account the teacher-student behavior trajectory modeling with relative and absolute attributes in space-time in the smart teaching environment, an attention recognition algorithm based on deep reinforcement learning is proposed to enhance the robustness and generalization ability of the attention recognition method, and to construct a space-time movement diagram that integrates the smart education data with the real classroom teacher-student interaction information in order to improve the accuracy of learning attention recognition. For the modeling and expression method of learning attention, a feature extraction and attention recognition method based on photoelectric volumetric pulse wave signal is proposed based on the personalized attributes of combining students' individual behavioral trajectories with wearable devices as the main means to realize the gated graph neural network fusion with enhanced spatio-temporal attributes.

3.2.3 Classification and hierarchical recognition model of learning emotions based on association detection

Based on the theory of cognitive load and the theory of emotional cycle of learning process, we carry out research on the classification and graded recognition of classroom learning emotions. Focusing on the study of teacher-student adaptive interaction in classroom teaching for attention and learning emotion recognition and collaborative analysis and enhancement technology, mainly including teaching object somatosensory part detection and recognition technology, teaching object multimodal data collaborative understanding, multi-view classroom teaching object target matching, adaptive assisted education service technology, establishing a classroom group relationship network with teacher-student teaching as the starting point, so as to build a CNN-based classification and grading recognition model of learning emotion.

Based on the learning analysis framework, we study the characteristics and connotations of intelligent teaching activities, refine the classification framework of teaching context data, classify intelligent teaching context data from both direct and indirect contexts and filter their specific indicators, focus on strong interactive tasks and task execution scenarios in intelligent teaching activities, build the foundation layer of the whole model, and realize the analysis of students' attention and learning emotion based on the learning analysis framework Extraction.

In this paper, we propose a repetitive attention recognition method that correlates related contents between heterogeneous data, which realizes the correlation analysis of independent objects in complex image information and word tuples in text information in a fine-grained perspective to form an attention mapping graph of multimodal data, so as to improve the semantic understanding of multimodal data and realize the edge peeling of student group behavior analysis in complex educational networks. The method is based on classroom group relationship extraction by common event discovery, which can realize feature extraction and fusion among multimodal heterogeneous data, relying on specific multimodal data such as teacher behavior and student behavior to realize the behavior layer of the whole model, and can effectively extract the similar semantic relationships between the underlying features of different types of media data.

In this paper, we carry out research on learning confusion state detection based on multimodal data, take learning events in the process of smart teaching as feature elements, combine the problem of learning emotion recognition in complex logical learning environments, study the design and implementation of emotion elicitation, and explore the construction of an end-to-end multimodal data analysis method based on deep learning and the event layer of the hierarchical classification and recognition model of learning emotion.

The global group discovery algorithm for classroom teaching needs to anticipate the global information of the whole network in advance, which is difficult to adapt to the association structure with large scale and dynamic structural changes, and it cannot realize group discovery from a specific node. This paper addresses the problem of differences in the structure and type information of different nodes in heterogeneous information networks and proposes a student group discovery algorithm based on digitally signed association link detection, which constructs a model activity layer by taking the teaching activities of association-like network nodes as reference objects and realizes classroom group

ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

discovery by fusing base vectors and signature vectors.

3.2.4 Classroom Behavior Prediction and Assisted Instruction System

To construct a behavior prediction and assisted teaching system based on collaborative modeling of multimodal data of teaching objects. By collecting log data from the intelligent guidance system, a two-layer Hidden Markov Model is used to model sequential learning behaviors using the research paradigm of educational data mining; after completing the mapping of data, a character-specific trajectory model is constructed by including location data in the communication domain data to realize the representation of character-specific behaviors in the combination of cyberspace and physical space. Combined with classroom behavioral event prediction that incorporates trajectory environment features, the captured learning behavior patterns are compared to describe effective learning behavior patterns and build an assisted teaching system. The specific architecture is shown in Figure 3.

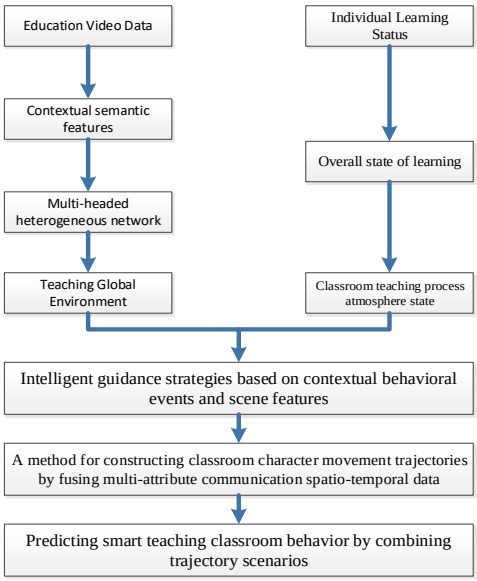

Figure 3: Classroom Behavior Prediction and Assisted Instruction System

To address the difficulty that the current classroom teaching quality assessment mainly relies on the subjective judgment of professionals, which is time-consuming and cannot be scaled up on a large scale, the study integrates classroom behavior, teacher-student emotion and speech for automatic teaching assessment, and achieves localization in local context and extraction at local scene feature analysis. Unstructured classroom instruction contains a variety of information, not all of which contribute to classroom behavior prediction. To address this challenge, this paper develops two attention-based multi-headed heterogeneous networks that incorporate event representations and their environments for classroom-level prediction at both local and global levels, obtains the global environment by adaptively aggregating the local environment of the environment, employs converters as core algorithm modules to encode contextual events and environments, and uses a multi-headed heterogeneous network to retrieve the overall environment of instruction. Parameters such as students' overall participation in teaching activities, students' average attention and average effective participation time are used as the basic characterization attributes of the classroom teaching process atmosphere state, and parameters such as individual attention, effective participation time, interaction time and frequency are used to characterize individual students' learning state, and the overall and individual learning state data are integrated to evaluate the classroom teaching process atmosphere state, and output the classroom teaching process atmosphere state based on contextual behavioral events and This paper proposes an intelligent learning guidance strategy based on contextual behavioral events and scene features.

In this paper, we propose a method for constructing classroom character movement trajectories by fusing multi-attribute communication spatio-temporal data. First, after preprocessing operations on classroom video and log data, the absolute spatio-temporal information and relative spatio-temporal information are fused into a behavioral spatio-temporal movement graph, and the edges of the graph are associated with neighboring and non-neighboring check-in records to strengthen the transfer relationship between different character locations in the classroom; second, the update of trajectory nodes is realized by constructing a gated graph neural network to achieve accurate teacher and student preferences for different location capture. This trajectory construction method achieves comprehensive

ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

coverage of multiple attributes of spatio-temporal data and representation of deep-level user physical spatio-temporal features, which further improves the accuracy of simulated character location preferences.

Finally, by modeling the behavioral data of teachers and students, the classroom teaching situation of teaching the same content several times is compared and analyzed to find out the differences between the two teaching sessions, and based on the analysis results, the classroom behavior prediction of smart teaching combined with trajectory scenarios is realized to provide a basis for the development evaluation of smart teaching.

4. Experiment

All experiments were conducted on a Lenovo workstation equipped with intel I9-10900k, 64GB RAM, 2*RTX3090, 1T pcie3.0 SSD. All experiments were conducted on the tensorflow platform.

First, the classroom learning affective state detection scheme was determined, and the classroom learning process was designed from two independent variables: the way of learning resources presentation and the presence or absence of online tests for learning activities. Secondly, 30 (total 120) valid subjects were selected each, the multimodal synergistic data of learners were collected, the micro-expressions, micro-motions, and physiological signals of learners were focused on analysis and association calculation, the prediction model was fitted, the classification and hierarchical recognition model of learning emotions based on association detection was designed and built, and then experimental tests were conducted, and if the new subjects exceeded the prediction trend range of the model, the emotion state prediction model is corrected. Finally, the experimental data and subjective interview data were combined to analyze the degree of influence of the presentation of resources in the teaching process and the provision of guided learning assistance in teaching activities on learners' affective states, and to propose corresponding guided learning strategies and support services. The specific experimental results are shown in Table 1.

Table 1: Experimental results

No.	Model	Top-1	Top-5	Fluid inference time(ms)	TensorRT inference time(ms)
1	AlexNet[1]	56.72%	79.17%	3.083	2.566
2	GoogLeNet[2]	70.70%	89.66%	6.528	2.919
3	SqueezeNet1_0[3]	59.60%	81.66%	2.74	1.719
4	VGG11[4]	69.28%	89.09%	8.223	6.619
5	MobileNetV1_x0_25[5]	51.43%	75.46%	2.283	0.838
6	MobileNetV2_x0_25[6]	53.21%	76.52%	4.267	2.791
7	ShuffleNetV2[7]	68.80%	88.45%	6.101	3.616
8	ResNet34[8]	74.57%	92.14%	5.668	3.424
9	ResNeXt101_64x4d[9]	78.43%	94.13%	41.073	31.288
10	DarkNet53[10]	78.04%	94.05%	11.969	6.3
11	Xception41[11]	79.30%	94.53%	13.757	7.885
12	InceptionV4[12]	80.77%	95.26%	32.413	17.728
13	**Method of our paper**	84%	**96.88%**	138.67	51.059

Based on the training methods of conditioning hyperparameter training and data augmentation methods used in the previous section, the network structure is adjusted, and the models are inferred on the three datasets Fer2013[13], CK+[14], and MMI[15] after labeling, and the accuracy is compared with the ground truth of the datasets to obtain the accuracy, and the model with the highest accuracy is finally selected.

ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

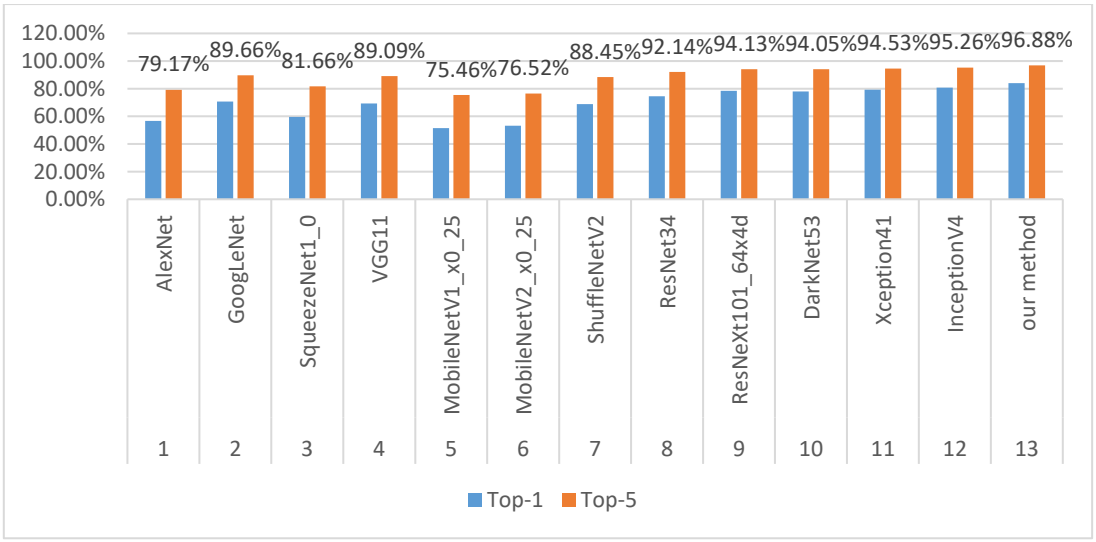

Figure 4: Comparison of recognition rate in the case of TOP-1 and TOP-5

From the Figure 4, the method proposed in this paper achieves 96.88% accuracy in practical application.

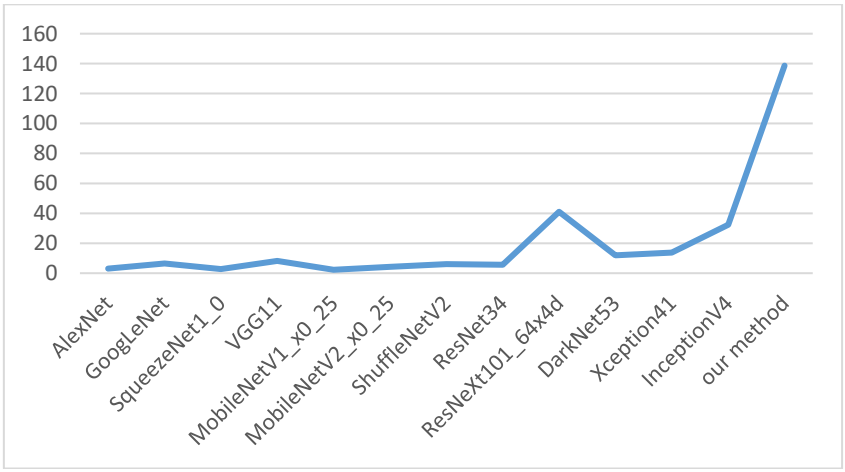

Figure 5: Fluid inference time (ms)

From Figure 5 and 6, it can be seen that the performance of the structure proposed in this paper is not good in both Fluid inference time and TensorRT inference time, which is mainly caused by the increase in the number of convolutional and hidden layers to improve the recognition rate.

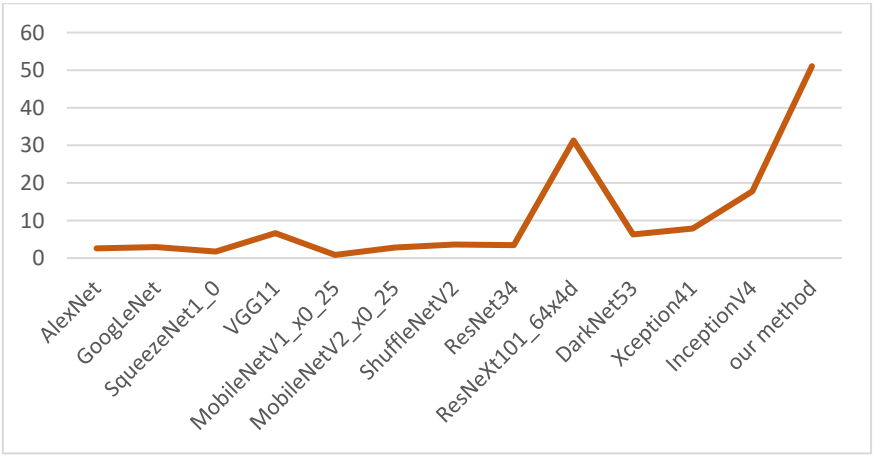

Figure 6: TensorRT inference time (ms)

ISSN 2522-6398 Vol. 6, Issue 12: 119-129, DOI: 10.25236/FER.2023.061222

5. Conclusion

This project focuses on classroom and online learning behavior analysis and key technology research based on emotion computing theory. Based on self-built expression dataset and deep network structure improvement, we break the bottleneck of small size and unbalanced data types of face expression dataset in the existing learner emotion computing process. Based on migration learning to enhance the face expression dataset, we solve the problem that the face expression dataset is quite difficult to obtain in the process of learner emotion calculation. Based on classification optimization, the problem of missing expression definition by CNN model in the current learner sentiment calculation process is solved. By building a single-model multi-task deep neural network, the problem that it is difficult to achieve real-time multi-tasking in the actual face monitoring scenario during learner emotion computation is solved. Finally, by building a classroom education simulation environment and comparing with the current mainstream CNN-like algorithms, the recognition rate section reaches 96.88%, but the overall computation time exceeds the average algorithm time by about 45%. Next, the whole research will focus on the improvement of the model structure, and it is proposed to achieve the optimization of the algorithm computation time through the form of pipeline.

References

[1] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.
[2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409. 1556, 2014.
[3] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. arXiv preprint arXiv:1704.04861, 2017.
[4] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.
[5] Koonce B, Koonce B. MobileNetV3 [J]. Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization, 2021: 125-144.
[6] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.
[7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
[8] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks [C] //Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.
[9] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
[10] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C] // International conference on machine learning. PMLR, 2019: 6105-6114.
[11] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.
[12] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
[13] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).
[14] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C] // International conference on machine learning. PMLR, 2019: 6105-6114.
[15] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. Advances in neural information processing systems, 2017, 30.