International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 5477-5488

https://doi.org/10.70517/ijhsa463473

# Research on Affective Computing and Multimodal Learning Behavior Data Mining in Civic Education

**Zhenyu Zhan[1,*] and Haining Wang[1]**
[1] School of Marxism, Shandong University, Jinan, Shandong, 250000, China
Corresponding authors: (e-mail: 202220141042@mail.sdu.edu.cn).

**Abstract** With the development of information technology, the traditional mode of Civic and Political Education has become insufficient. This study explores the application value of affective computing and multimodal learning behavior data mining in Civic and Political Education. A multimodal classroom environment for Civic and Political Education was constructed based on constructivist learning theory and multimodal discourse analysis theory, and a multimodal learning behavior analysis model integrating deep learning and Bayesian network was designed. The experiments used hybrid discriminant restricted Boltzmann machine (HDRBM) neural network to process the multimodal data, and analyzed the learning causality through Bayesian network. The study invited 30 college undergraduates to participate in the experiment, and the results showed that the percentage of individual students' focused emotions identified by the system was 47.8%, which was close to the 54.2% of the manual statistics; in the analysis of the overall students' emotions, the focused emotions identified by the system was 46.78%, and the manual statistics was 51.42%, and the errors of both of them were small. The frequency analysis of multimodal behaviors shows that the frequency of A7 (teacher-oriented) behaviors in students' participatory learning is the highest; in focused learning behaviors, the ratio of students' gaze on learning aids (R4) gradually tends to 1 from more than 1 at the beginning of the semester, which indicates that students' focus on the classroom gradually increases. The study proves that multimodal sentiment analysis and learning behavior mining can effectively improve the teaching effect of Civics education and provide new ideas for Civics education innovation.

**Index Terms** Affective computing, Civic education, Multimodal data mining, Deep learning, Learning behavior analysis, Bayesian network

## I. Introduction

Civic and political education is an education with distinctive Chinese characteristics, and its importance as a key link in shaping young people's thoughts and leading the social trend to understand the socialist core values more deeply, so as to better serve China's socialist construction is self-evident. With the acceleration of globalization and the development of information technology, youth groups are facing unprecedented ideological impacts and value choices [1], [2]. In this context, the traditional teaching mode of Civic and Political Education, student participation, teacher-student interaction, teaching evaluation and so on, has exposed defects. And nowadays, vigorously integrating science and technology and education to keep up with the pace of the times is an important process of developing and innovating new forms of education [3]. With the help of information technology, exploring students' learning behavior and emotional feedback provides a path for the high-quality development of ideological education.

The concept of affective computing was proposed in 1997 by Prof. Picard of MIT Media Laboratory, who pointed out that affective computing is a computation that is related to, derived from, or able to exert influence on emotions [4]. It technology is an emerging field that integrates psychology, computer science and education. By analyzing large amounts of text, speech and image data, affective computing technologies can extract and understand emotional information from it and further classify and predict emotions [5]. The potential of such technologies is widely used in educational environments and can lead to a more personalized and efficient learning experience. During the teaching and learning process, the impact of students' affective states on learning outcomes cannot be ignored. Several studies have shown that positive affective states increase students' motivation and engagement, while negative affect can lead to problems such as distraction and learning disabilities [6]-[9].

By utilizing sensors, artificial intelligence, and machine learning, affective computing technology is able to collect students' emotional data in real time, including physiological signals such as facial expression, voice pitch, heart rate, etc., to recognize, interpret, and simulate human emotional expressions, so as to obtain detailed emotional analysis [10]-[13]. Moreover, personalized learning is precisely one of the outstanding potentials that affective computing has shown in education. By analyzing students' emotions in the classroom teaching environment, educators can make adjustments to each student's needs, enhance students' learning interests, and encourage students to continue in-depth learning when they show high emotions, which is a dynamic feedback mechanism that not only improves the learning efficiency, but also enhances students' sense of participation [14]-[16]. With the continuous advancement of information technology, affective computing may become a key driver in shaping the future of education. Student behavioral data mining provides educational institutions with a more comprehensive picture of students by analyzing their learning status, learning behavior process, etc., so as to better guide students' learning [17]. In the field of civic education, single-modal mining has matured, but multimodal fusion is still in its infancy. In order to deepen the Civic Education and improve the teaching effect, the behavioral data analysis of multimodal is of great value.

Civic education, as a key link in shaping young people's thoughts and leading social trends, is of great significance for a deeper understanding of socialist core values and serving the construction of socialism in China. With the accelerated process of globalization and the rapid development of information technology, youth groups are facing unprecedented ideological impacts and value choices. The traditional model of Civic and Political Education has revealed obvious deficiencies in student participation, teacher-student interaction, and teaching evaluation, making it difficult to meet the needs of Civic and Political Education in the new era. Nowadays, the field of education is actively integrating science and technology with teaching to keep pace with the times, which has become an inevitable trend of innovative forms of education. It is particularly important to explore students' learning behavior and emotional feedback with the help of information technology to provide a new path for the high-quality development of Civic and Political Education. The concept of affective computing was proposed by Prof. Picard of MIT Media Lab in 1997, which is a computation related to, derived from, or capable of influencing emotions, integrating the emerging fields of psychology, computer science, and education. By analyzing large amounts of text, speech, and image data, affective computing techniques can extract and understand emotional information and perform emotion classification and prediction. Several studies have shown that positive affective states increase student motivation and engagement, while negative affect can lead to distraction, learning disabilities, and other problems. Affective computing technology utilizes sensors, artificial intelligence and machine learning to collect student emotion data in real time, including physiological signals such as facial expression, voice pitch, heart rate, etc., to identify, interpret and simulate human emotion expression and obtain detailed emotion analysis. By analyzing students' emotions in the classroom environment, educators can make adjustments to each student's needs, enhance learning interest, and form a dynamic feedback mechanism that not only improves learning efficiency, but also enhances students' sense of engagement. Student behavioral data mining provides educational institutions with a more comprehensive picture of the students' situation through the analysis of the learning status and the learning behavior process, so as to better guide learning. In the field of civic education, single-modal mining has become mature, but multimodal fusion is still in its infancy. In order to deepen the Civic and Political Education and improve the teaching effect, multimodal behavioral data analysis has important research value.

Based on the above background, this study constructs a multimodal classroom environment for Civic and Political Education, designs a multimodal learning behavior data analysis model based on deep learning and a Bayesian network learning causality analysis method. The experimental part of the study analyzes students' emotional state and learning behaviors in the learning process of Civic and Political Science courses in multiple dimensions, verifies the validity of the model through the comparison of system identification and manual statistics, and analyzes the frequency of learning behaviors from the three dimensions of participation, initiative, and concentration, so as to provide data support and theoretical guidance for the reform of Civic and Political Science education and teaching.

## II. Multi-modal Civics Classroom Environment Construction

### *II. A. Constructing ideas*

The construction of the multimodal classroom environment for ideological and political education is guided by the constructivist learning theory, combined with the multimodal discourse analysis theory, especially the unique collaborative win-win effect of multimodal cognition, to explore the convergence points between the constructivist learning theory, the multimodal discourse analysis theory and the construction of the multimodal classroom environment for ideological and political education. The aim is to change the phenomenon of "teachers having extreme say" and students having "extreme lack of say" existing in most ideological and political education classrooms in colleges and universities, and to provide new ideas for the teaching reform of ideological and political education classrooms. The multimodal classroom environment emphasizes the use of multimodal teaching courseware as a tool, utilizes multimodal interlaced and personalized teaching methods, and through diversified teacher-student communication methods and multimodal learning materials and classroom activities, supplemented by a diversified classroom assessment system, fully mobilizes students' enthusiasm and initiative in classroom participation, encourages students to experience the learning process and feel the learning outcomes. Gain a sense of achievement in learning, and thereby improve one's own learning ability and comprehensive quality ability.

## *II. B.Multimodal Learning Behavior Analysis Techniques*

### II. B. 1) Multimodal data

Modality is an objectively existing symbolic system that can be interpreted in terms of specific perceptual processes, such as speech, image, action, expression, gesture, etc. can be called a modality after transcoding [18]. Through the analysis of modality and data mining, the development state of things can be dynamically measured, so as to make corresponding predictions and judgments. However, due to the diversity of natural environments and events, especially the wide application of information technology in the field of education, which makes it possible for the applicant to obtain information from a variety of channels, multimodal data is thus created. Multimodal refers to a symbolic system that integrates the use of several modalities in analyzing an event, thus enabling interaction between human sensory organs and the external environment. Multimodal information is very complex and involves a variety of symbolic representation systems, such as visual, auditory, olfactory, gustatory, tactile, etc., which are usually not used singly but in combination. For example, an online learning platform may involve learners' audio data, video data, EEG data, eye movement data, action gestures and other data, which come from different sources and forms, and combine with each other to form a multimodal data environment. An important feature of multimodal data is that it is complementary; data from one modality can often only explain the process of an event from one aspect, and this information cannot be obtained from other modalities. Multimodal data can fuse data such as voice, text, video, and physiological information to achieve a more comprehensive, scientific, and systematic judgment and understanding of things.

### II. B. 2) Multimodal learning behavior data

The science and validity of learning data is the key to multimodal learning analysis. Blended learning mode is based on online and offline learning modes such as microteaching, catechism, SPOC, etc. Through teacher-student interaction, student-student interaction, teacher-student interaction and teacher-student interaction with teaching equipments, learners can understand and master the learning content in activities such as online learning, group discussion and collaborative learning. According to the information characteristics in the learning data, the multimodal learning analysis data of blended learning is divided into five categories: basic information data, learning signs data, human-computer interaction data, learning resources data and learning context data. Among them, the basic information data refers to the demographic data such as gender, age, and grade of learners, which is a necessary data set for all learning analytics. Learning signs data include physical movement behavior data and physiological behavior data such as brain and heart responses. Human-computer interaction data includes mobile interface data such as fingerprints and iris and natural language data such as dialog and discussion. Learning resource data includes online resource data such as sound and video and offline resource data such as text and multimedia courseware. Learning context data include physical space data such as weather and temperature and virtual space data such as socialization and interaction [19].

### II. B. 3) Learning behavior data analysis process

The multimodal learning analysis process of blended learning is shown in Figure 1.

(1) Use data collection devices (e.g., cameras, eye-tracking devices, smart wearable devices, etc.) to collect multimodal data such as teachers, learners, teaching environment, learning resources, etc. from online teaching platforms, teaching software and offline classrooms, experimental internships, and other teaching environments, as the data basis for multimodal data analysis.

(2) Perform data cleaning and extraction on the massive multimodal data collected, delete invalid data, select key attributes for data fusion, and determine the relevance and credibility of the data to form a standardized and unified format of multimodal learning analysis data flow.

(3) On the basis of teaching theory and teaching experience, machine learning, data mining, statistical analysis and other technologies are used to predict and analyze the multimodal learning analysis data stream, excavate the teaching law, consider the teaching task, and ultimately form certain decision support and recommendations.

(4) The decision report formed by data analysis is visualized and recommended to teachers, learners and teaching managers. Teachers can know the learning situation of the whole class based on the decision report, adjust the teaching progress and teaching methods in time, and give personalized learning plans to learners with different learning levels. Learners can also understand their own learning status clearly and comprehensively, find their shortcomings in time and make up for the shortcomings. Teaching managers can grasp the learners' thought dynamics, learning mood changes, etc., in order to better serve the learners. At the same time, the school can also

make macro decisions based on the decision report, optimize the education and teaching mode, and promote the benign development of the education ecosystem.

(5) The results of the data decision report are further fed back into the online and offline teaching environment to realize the analysis and optimization of the whole process and all-round teaching.

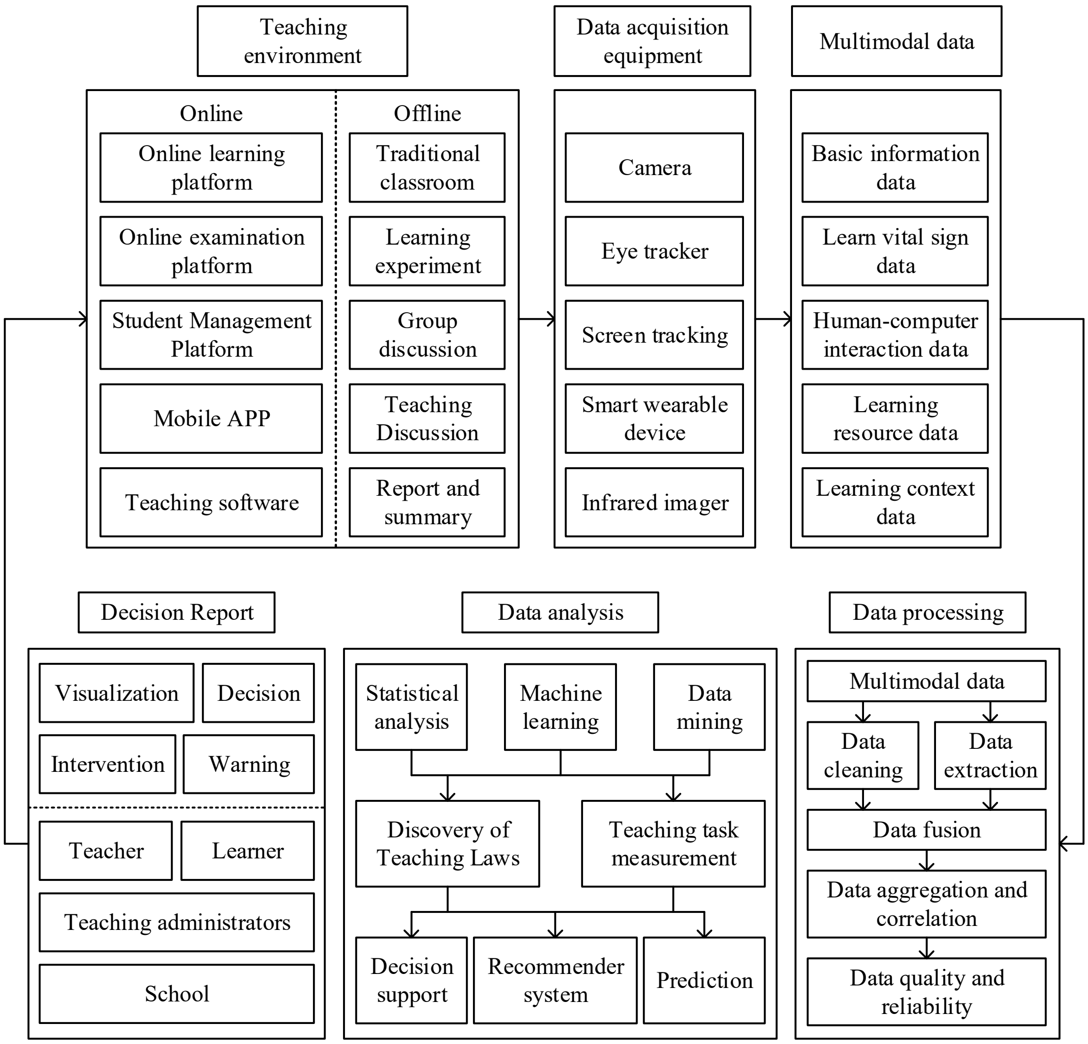



Figure 1: Shows the multimodal learning analysis process of hybrid learning

## III. Deep learning based multimodal learning behavior data analysis model

Learning behavior analysis is a complex and high-dimensional analysis process, and traditional analysis methods are difficult to analyze multimodal data, and the emergence of big data and intelligent algorithms provides new methods for student learning characterization and educational big data mining. This paper proposes the use of deep learning algorithms for multimodal learning analysis, which provides a new paradigm for mining learner behavioral characteristics using multimodal educational big data in the field of educational technology.

### *III. A. Design of Learning Behavior Analysis Model Supported by Deep Learning*

Deep learning is to learn the intrinsic laws and representation levels of the sample data, and the information obtained in these learning processes can be of great help in the interpretation of data such as text, image and sound. Construct a big data acquisition framework for multimodal learning behavior data, collect real-time time-series multimodal data such as classroom, laboratory and teaching platform, clean and preprocess the multimodal learning behavior data according to the modal characteristics of the data and the characteristics of the factors, and then use intelligent algorithms to perform shallow feature analysis, and then normalize them and deep fusion [20]. Based on the high-dimensional data processing needs to construct a novel streaming deep learning analysis model to analyze the deeply fused multimodal data. Given that the deep hybrid discriminative restricted Boltzmann machine (HDRBM) neural network has the advantages of dimensionality reduction, classification, regression, collaborative filtering, feature learning, and topic modeling, and collects the advantages of feature extraction, high accuracy, and speed of the current mainstream deep learning, it is the most suitable deep learning framework for processing multimodal data at present. Therefore, in this study, HDRBM neural network is used to process multimodal learning behavior data in the smart learning environment, in order to establish a multimodal learning behavior analysis model and to characterize learner behavior.

### *III. B. Learning behavior feature extraction based on HDRBM*

After the multimodal data is obtained by obtaining the same mapping for a number of deep learning networks, in order to further extract high-dimensional features based on this mapping, meters are generated using a combination of discriminative constrained Boltzmann machines and bull-forming constrained Boltzmann machines, i.e., HDRBM [21]. The generative model can solve the problem of a small number of training datasets and is better than the discriminative model used in the classification problem in the case of a small number of datasets, but as the multimodal data continues to increase, the discriminative model is better than the generative model when a large amount of training data is available. In order to take into account the advantages of both, this study proposes to compose a hybrid generative constrained Boltzmann machine and a discriminative constrained Boltzmann machine to form a hybrid discriminative constrained Boltzmann machine in multimodal deep learning modeling, and design the corresponding multimodal neural network objective function and optimization pathway accordingly.

For training set . $D_{tain}$ ., the Boltzmann machine is generated to minimize the negative log-likelihood as the objective function, i.e., the minimization objective function:

$$L_g(D_{train}) = -\sum_{i=1}^{|Dtrain|} \log p(x_i, y_i) \tag{1}$$

where $p(x_i, y_i)$ is the joint distribution of a sample $x$ and a label $y$.

The generative Boltzmann machine can obtain the joint probability distribution between learning behavior characteristics and learning evaluation. The discriminative Boltzmann machine is able to accurately label the input learning behavior features for learning evaluation. Consider negative log-likelihood with an objective function:

$$L_d(D_{train}) = -\sum_{i=1}^{|Dtrain|} \log p(y_i \mid x_i) \tag{2}$$

In order to adapt to the case of low data volume in the early stage and also to meet the performance requirements in the case of sufficient data set in the later stage, combining the objective functions of Eq. (1) and Eq. (2), the objective function of the hybrid discriminative constrained Boltzmann machine is:

$$L_h(D_{train}) = L_d(D_{train}) + \alpha L_g(D_{train}) \tag{3}$$

where the adjustable parameter $\alpha$ indicates the weight of the generating model on the overall model. If $\alpha$ is large, it favors generating Boltzmann machines. Conversely, it favors discriminative Boltzmann machines. For the selection of . $\alpha$ . in the experiment, we draw on the e-greedy strategy, which is mature and widely used in reinforcement learning (RL), so that the hybrid discriminative restricted Boltzmann machine can automatically reduce the influence weight of the generative model on the whole model when the dataset grows, thus obtaining a more flexible and accurate model effect.

On the other hand, during Boltzmann machine training, gradient instability scenarios, such as the gradient dispersion and gradient explosion problems, occur as the number of network layers increases, and either disappear or explode in hidden layers close to the input layer. For this reason, we invoke the batch normalization (BN) method to circumvent the gradient dispersion problem. It is applied to each layer before the activation function, which is to do mean and variance normalization, and for each batch of data also make zoom in, zoom out, and panning, in order to converge faster when the gradient decreases. In addition, in the case of a small number of datasets and a large number of network layers in the early stage, it is easy to make the whole model overfitting phenomenon, so the Dropout strategy commonly used in neural networks is introduced to improve the generalization performance of the model. During the training process, the Dropout strategy randomly sets a part of a layer of units to 0, which is equivalent to discarding the connection between the unit and the neuron in the next layer, so as to reduce the probability of model overfitting.

### *III. C. Learning causality analysis based on Bayesian networks*

#### III. C. 1) Bayesian network structure learning

In order to mine the causal relationship between learning performance and multimodal data features, a two-layer Bayesian network structure is constructed. Bayesian network structure learning is to analyze the sample dataset to obtain the a priori information, and then to find the interrelationships among the nodes and build the corresponding network structure. The learning of Bayesian network structure is mainly divided into two main categories. They are the methods based on scoring search and the methods based on dependent statistical analysis, among which the former process is simple and standardized, so it is more commonly used [22]. The scoring function is used to evaluate how well the network topology fits the sample set dataset, and the scoring function will be based on a blank network, traversing all the nodes in front of the node to be solved according to the order of the nodes, and subsequently selecting the node with the largest a posteriori probability value as the parent of the node. It is then connected by directed edges and traversed until the network is constructed. The commonly used scoring function is shown in equation (4):

$$\log P(DIG) = \sum_{i=1}^{n}\sum_{j=1}^{q_i}\left[\log(\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+N_{ij})}) + \sum_{k=1}^{x}\log(\frac{\Gamma(\alpha_{ijk}+N_{ijk})}{\Gamma(\alpha_{ijk})})\right] \tag{4}$$

In the above equation, $D$ is the dataset, $G$ is the topology, $n$ is the number of nodes, $\Gamma(\alpha_{ij})$ is the gamma function, $r_i$ denotes the number of states of the node $X_i$, $\alpha_{ij} = \sum_k \alpha_{ijk}$, and $q_i$ denotes all combinations of values taken by the parent of node $X_i$. $N_{ijk}$ denotes the number of samples corresponding to a parent node state value of $j$ when the node value is $k$.

#### III. C. 2) Bayesian network parameter learning

After learning the topology of the Bayesian network through the structure learning algorithm, the next step is to use the parameter learning algorithm to learn the parameters of the Bayesian network, that is, to find out the conditional probability distribution of the variables with respect to the parent node, as the a priori information for posterior inference. If the variables obey the Bayesian distribution, the Bayesian parameter prior distribution is as follows:

$$\begin{gathered}\hat{\theta}_{ijk} = \frac{a_{ijk}+N_{ijk}}{\Pi_k(a_{ijk}+N_{ijk})} = \frac{a_{ijk}+N_{ijk}}{a_{ij}+N_{ij}} \\ \forall \mathrm{i} \in [1:\mathrm{n}].\forall \mathrm{j} \in [1:\mathrm{q_i}].\forall \mathrm{k} \in [1:\mathrm{r_i}]\end{gathered} \tag{5}$$

The posterior distribution of the parameters is as follows:

$$\begin{gathered}P(\theta_{ij}/G,D) = \frac{P(\theta_{ij}/G)P(D/G,\theta_{ij})}{P(D)} = \frac{\Gamma(a_{ij}+N_{ij})}{\Pi_k\Gamma(a_{ijk}+N_{ijk})} \\ \Pi_k\theta_{ijk}^{a_{ijk}+N_{ijk}} = Dir(a_{ij1}+N_{ij1}, a_{ij2}+N_{ij2}, \ldots, a_{ijr_i}+N_{ijr_i})\end{gathered} \tag{6}$$

Then the maximum a posteriori probability of the parameter θ is estimated as:

$$\begin{gathered}\hat{\theta}_{ijk} = \frac{a_{ijk}+N_{ijk}}{\Pi_k(a_{ijk}+N_{ijk})} = \frac{a_{ijk}+N_{ijk}}{a_{ij}+N_{ij}} \\ \forall \mathrm{i} \in [1:\mathrm{n}].\forall \mathrm{j} \in [1:\mathrm{q_i}].\forall \mathrm{k} \in [1:\mathrm{r_i}]\end{gathered} \tag{7}$$

The network parameters are learned by Bayesian formulation, which organically combines the a priori information and the sample dataset D, and effectively improves the accuracy of parameter learning. The vector $V = (v_1, v_2, \ldots, v_5)$ denotes the set of variables, . $v_i$ . denotes the mixed data features, physiological data features, psychological data features, behavioral data features, and learning context elements, in that order, and n sets of observable datasets are selected $X = \{x_1, x_2, \ldots, x_n\}$ are trained to derive the causal relationship between the variable V and the evaluation of learning behavior. After the Bayesian model training using the constraint-based method, the causal network of high-dimensional variables can be constructed, i.e., the causal relationship-based Bayesian network model is shown in Fig. 2.

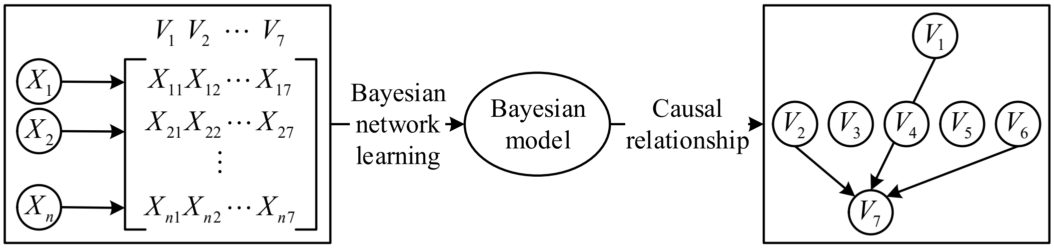



Figure 2: Causal analysis process based on Bayesian network

## IV. Multimodal Emotional Analysis of Civics Classroom

### IV. A. Experimental design

In this study, 30 volunteers were invited to carry out the experiment, they are all undergraduate students from a university, we selected a ten-minute video of the Civics course to watch and learn, and finally captured the learning video of 30 learners. After the subjects watched the video, they were invited to conduct a survey on the emotional changes in each time period during the viewing process, that is, they viewed the recorded learning process, and made a judgment on the learning emotions expressed in each minute. Since the subjects were more concerned about their real emotions hidden under their facial expressions and body postures, we also asked the experimenters to judge only from the image level, and if the two judgments were consistent, the emotions were recorded directly for that time period. If they agree, the type of emotion learned at that time will be recorded directly. If there is a disagreement between the two, a second experimenter will make the judgment and confirm the final emotion type.

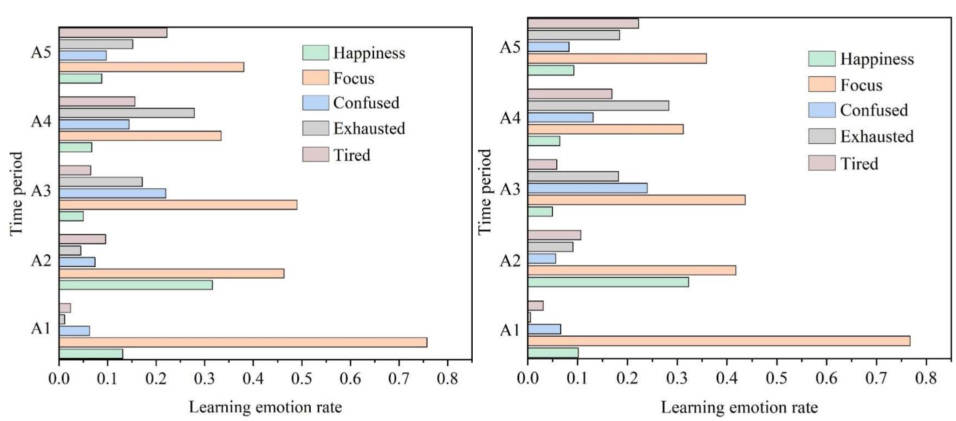
Considering that the amount of work would be huge if the manual recognition statistics of each image were performed for each learner, one subject Q will be randomly selected in this study to perform the manual statistics of emotions for each image of the video intercepts, which will be used for the comparative analysis of the results of individual emotion recognition. The other subjects will directly record the main types of emotions for each time period by watching the recorded learning process, and finally the manually counted individual learning emotion data will be summarized as a whole for the comparative analysis of the overall emotion recognition results.

Finally, after the manual statistics are completed, the system is used to record the learning emotion recognition of the acquired information of the students' learning process, and the emotion recognition results of the subjects Q and the whole are taken out for the comparative analysis of the data with the emotion results of the manual statistics. If the recognition accuracy of the system recognition results in each emotion type is relatively high, it means that the system's learning emotion recognition results are reliable, thus proving the effectiveness of the system.

### IV. B. Results of Sentiment Analysis of Multimodal Learning Behaviors

#### IV. B. 1) Individual Student Emotional Analysis

In this study, the results of the experiments will be analyzed and compared in terms of both the individual emotion recognition results and the overall emotion recognition results · so as to make a judgment on the effectiveness of the system. The recognition results are shown in Figure 3. Where A1-A5 represents the time period.




(a) Individual system recognition data for the participants q

(b)Individual statistics of individual data of the subjects q

Figure 3: Test and test of individual emotional analysis of students

Figure 3a shows the multimodal emotion recognition system statistics of subject Q in the various time periods of learning emotion change data, from the figure can be seen that subject Q in ten minutes of emotion change is small, basically in the state of focus on listening and learning, but the emotional state in the second time after the second time period into the low tide, produce more obvious boredom and doubt situation. Figure 3b shows the results of the emotional data from the manual statistics of subject Q. From the figure, it can be seen that the results of the manual statistics avoided some smaller bases, and the manual statistics will have a greater degree of differentiation in the identification of the two states of boredom and confusion than the system identification. Whether it is the result of system identification or manual statistics, the overall change of learning emotion of subject Q is in the state of focusing on listening, and with the increase of learning time tired emotion gradually rises, and focusing emotion tends to stabilize the state.

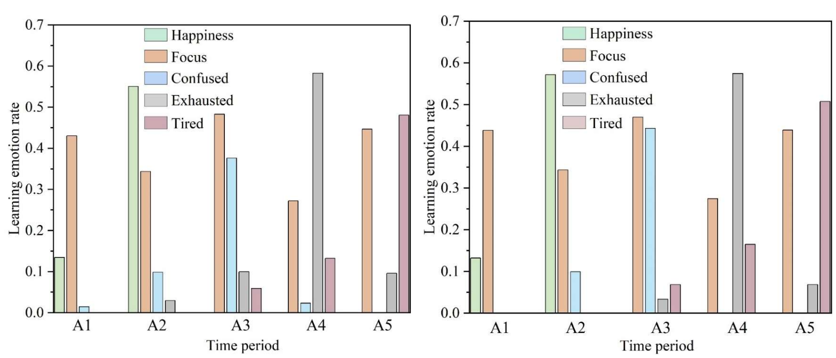
In order to better compare and analyze the experimental data, different types of learning emotion data of all time periods are extracted for analysis and comparison, and the results are shown in Table 1, which shows that for each type of emotion data, although there is a certain degree of error in the manual statistical results and the system recognition results, the error value is within a small range, and there is no big deviation. Therefore, it can be proved that the multimodal emotion recognition system designed in this study is effective in individual emotion recognition. It is proved that the multimodal emotion recognition system proposed in this study is effective for the results of the performance of a single individual student in an online learning classroom.

Table 1: system statistics and artificial statistics are compared

| | Amused | Delicated | Confused | Exhausted | Tired |
|---|---|---|---|---|---|
| System recognition | 13.34% | 47.8% | 12.69% | 13.8% | 12.37% |
| Manual statistics | 11.9% | 54.2% | 8.32% | 10.36% | 15.22% |

**IV. B. 2) Overall student sentiment analysis**

The multimodal sentiment analysis system can not only identify and analyze the sentiment of individual learners, but also identify and analyze the sentiment state of all students studying the course, and the identification results are shown in Figure 4. Figure 4a shows the overall results of the multimodal emotion recognition system on the 30 subjects studying the course, from which it can be seen that in the fourth period of time the students' learning state began to decline, boredom and tiredness became more emotions, and then returned to the normal state. Figure 4b shows the graph of the overall learning emotion analysis results of the classroom after manual statistics, from the figure it can be seen that most of the students' learning emotions in different time periods are relatively similar, and only a small number of people show different learning emotions, which may be somewhat related to the subjects' individual learning interests, and it is not guaranteed that every subject has a high learning interest in the course.




(a)The participants overall system recognition data

(b) The overall artificial statistics of the subjects

Figure 4: The recognition of the overall emotional analysis of the learners

The manual statistical results of the overall learning emotion situation in the classroom and the system identification results are compared and analyzed, and the analysis results are shown in Table 2, which shows that comparing the manual statistical results of the same emotion labels with the system identification results reveals that there is also a certain degree of error, but the value of the error is relatively small. Therefore, it is proved that the multimodal sentiment analysis system proposed in this study is effective in recognizing the learning sentiment in the online classroom.

Table 2: The overall system statistics are compared to the artificial statistics

| | Amused | Delicated | Confused | Exhausted | Tired |
|---|---|---|---|---|---|
| System recognition | 13.34% | 46.78% | 10.68% | 16.12% | 25.71% |
| Manual statistics | 14.75% | 51.42% | 7.45% | 10.76% | 27.12% |

### *IV. C. Frequency analysis of multimodal behavior*

#### IV. C. 1) Participatory learning behaviors

Participatory learning behaviors are mainly reflected in the three modal dimensions of students' positional movement (M), body movement (A) in image data and technical support (T) in text data, which are the behavioral manifestations of students' “doing”. The observer and the researcher discussed and decided to collect 17 coded data in the three dimensions, counting the effective positional movement modal data (M1-M4) of students in the smart classroom, totaling 65 groups, the effective body movement modal data (A6-A10), totaling 281 groups, and the effective technical support for the textual data (T1-T8) derived from the student learning platform, resulting in a total of 130 groups of data. Due to the existing smart classroom environment equipment can not obtain invalid behavior and no behavior data, and human statistics have subjective differences causing collection difficulties. Therefore, no statistics were counted in this study, and the analysis of participatory learning behaviors was focused on effective participation in learning behaviors. According to the 3 dimensions of behavioral coding data to form a word cloud map to visualize the presentation as shown in Figure 5.

From the word cloud plots for each stage of the semester at the beginning, middle, and end of the semester, it can be seen that participatory learning behavior A7 (teacher-oriented) occupies the largest area of the word cloud in all of them, which indicates that this behavior occurs with the highest frequency. A6 and A9 behaviors are also more frequent, indicating that students have more behaviors such as viewing courseware, boards, and learning resources through the interactive whiteboard. In the smart classroom, students will use smart learning tools to realize combined online and offline learning activities. The learning report generated by the learning platform shows that T8 (information collection) and T5 (communication and discussion) behaviors appear more often, indicating that students can actively use learning tools to collect information and carry out cooperation and communication among cloud-based learning groups.

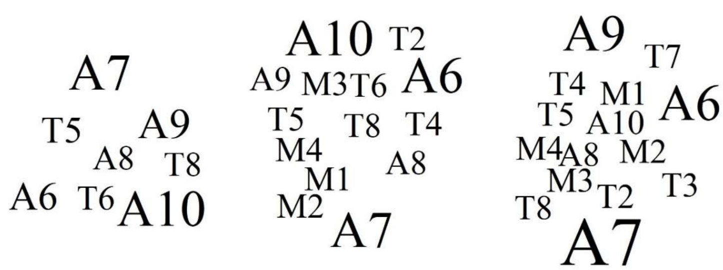



Figure 5: Word cloud

#### IV. C. 2) Active Learning Behavior

According to the qualitative characterization of learning behaviors in the multimodal coding table of students' learning behaviors, active learning behaviors are mostly embodied in language-based teaching and learning activities, and are the behavioral manifestation of students' learning “attitudes”, i.e., multimodal sound data (S). The sound data of active learning behaviors S1, S3, S4, S5, and S6 were obtained from the statistics of three semester classes at the beginning of the semester, the middle of the semester, and the end of the semester, and the frequency analysis of the five groups of data was performed to obtain the ratio of the number of active learning behaviors actually produced by students in a class to the total number of active behaviors that should be produced in that class. The closer the ratio is to 1, the higher the frequency of students' active learning behaviors. Visualizing the change in frequency, the ratio of the five groups of coded data shows an upward trend as the semester progresses and is constantly approaching 1. The results are shown in Figure 6, which shows that through the analysis of the frequency of students' active learning behaviors, it is clear that students' active learning behaviors that occurred less frequently at the beginning of the semester in the smart classroom gradually increased with the development of time, and that students' attitudes toward learning were effectively improved. For example, the active learning behaviors of students using learning tablets to raise their hands online, post results and evaluations, and actively answering questions and communicating and discussing offline increased.

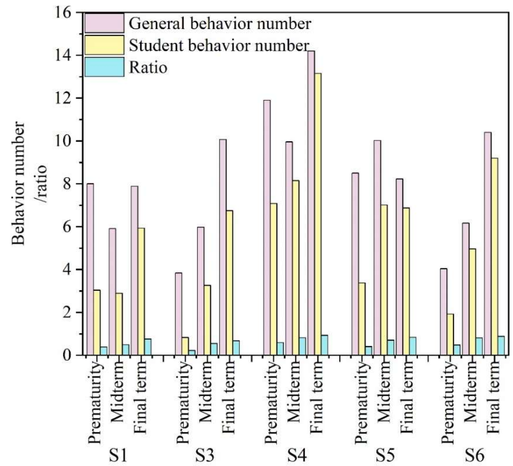



Figure 6: Initiative learning behavior analysis

#### IV. C. 3) Focused Learning Behavior

Concentrated learning behavior is mainly reflected in the physical actions in the image data (R1-R5). According to the definition and dimensions of concentration, learning concentration refers to a psychological state in which the learner focuses on the learning process and learning activities, emphasizing the durability of the time period and the stability of attention, which is a behavioral manifestation of the learner's learning “ability”. The observer counted the ratio of the duration of students' focused behavior in each learning activity to the total duration of learning activities in the classroom in seconds to determine the degree of students' focused behavior. The closer the ratio is to 1, the higher the student's concentration and learning ability. The results are shown in Figure 7, as the semester develops, students focus on the learning content, the number of teacher reminders decreases, students can focus on the interactive whiteboard in time (R1), students listen attentively to the teacher's lectures (R2), classmates' speeches (R3), and the concentration of watching the book content (R5) is gradually improving, and the concentration ratio gradually tends to be closer to 1. It was found that, at the beginning of the semester, the behavior of the students' gazing at the learning aids (R4) was found to be more than 1, indicating that the students were overusing the learning aids and their learning attention was shifted to the learning aids themselves. As the semester progresses, the frequency of this behavior decreases and converges to 1, indicating that the use of non-learning tools is decreasing and the students' attention to the classroom is gradually increasing.

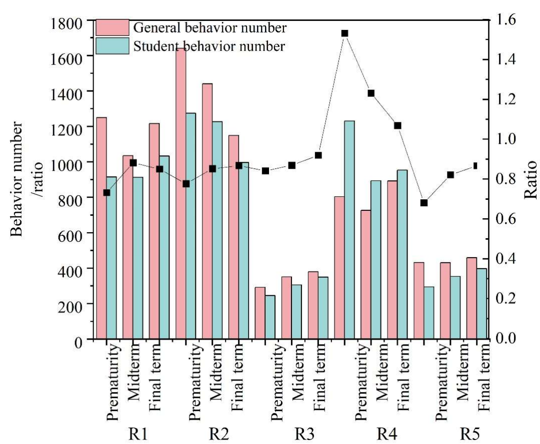



Figure 7: Focus on behavioral analysis

## V. Conclusion

In this study, we constructed a multimodal classroom environment for civic education, designed a multimodal learning behavior analysis model integrating deep learning and Bayesian network, and verified the application value of affective computing and multimodal learning behavior data mining in civic education through empirical research. The study shows that the designed multimodal sentiment analysis system can effectively identify students' learning sentiment states. In individual emotion analysis, the system's recognition rate of boredom emotion is 13.8%, while the manual statistics is 10.36%; the recognition rate of confused emotion is 12.69%, while the manual statistics is 8.32%, and the errors are all within the acceptable range. In terms of overall emotion analysis, the system's recognition rate of focused emotion for 30 subjects was 46.78%, which was similar to the 51.42% of manual statistics, proving that the system has high reliability in classroom emotion recognition.

The frequency analysis of multimodal learning behaviors found that among the participatory learning behaviors, the teacher-oriented (A7) behavior appeared with the highest frequency at all stages of the semester at the beginning, middle, and end of the semester, followed by watching the classroom materials through the interactive whiteboard (A6) and using the learning platform for data collection (T8) and communication and discussion (T5). The results of the analysis of active learning behaviors showed that the ratio of the five groups of coded data, such as students actively answering questions, raising hands online, publishing results and evaluation, showed an upward trend with the development of the semester, indicating that students' attitudes toward learning had been effectively improved. In terms of focused learning behaviors, the ratio of students' focus on the interactive whiteboard (R1), listening attentively to the teacher's lectures (R2), classmates' speeches (R3), and watching the book content (R5) gradually tends to be close to 1. In particular, it is noted that the ratio of students' gaze on learning aids (R4) behavior has gradually narrowed from more than 1 at the beginning of the semester to close to 1, which indicates that the use of unrelated learning aids has gradually been reduced, and the students' classroom focus has been significantly improved.

The research results provide a scientific multimodal emotion analysis and learning behavior mining method for Civic Education, which can help teachers understand students' emotional changes and learning status in a timely manner, adjust their teaching strategies, and realize personalized teaching, so as to improve the quality and effect of Civic Education.

## References

[1] Vakhshoozi, T., Sarookhani, B., & Sarabi, H. (2017). The effect of globalization on the moral values change among the youth. The Sociological Studies on Youth, 7(25), 2476-5198.

[2] Juraevna, K. T. (2020). Ideological and theoretical bases and problems of protection of young generation from the effects of globalization. International Journal on Integrated Education, 3(3), 67-69.

[3] Cloete, A. L. (2017). Technology and education: Challenges and opportunities. HTS: Theological Studies, 73(3), 1-7.

[4] Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., ... & Zhang, W. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. Information Fusion, 83, 19-52.

[5] Huang, Y., Fei, T., Kwan, M. P., Kang, Y., Li, J., Li, Y., ... & Bian, M. (2020). GIS-based emotional computing: A review of quantitative approaches to measure the emotion layer of human–environment relationships. ISPRS International Journal of Geo-Information, 9(9), 551.

[6] MacAllister, G. P. (2025). The Impact of Emotion-Based Teaching Strategies on Motivation in Collaborative Learning. Journal of Advanced Research in Education, 4(1), 7-10.

[7] Denovan, A., Dagnall, N., Macaskill, A., & Papageorgiou, K. (2020). Future time perspective, positive emotions and student engagement: a longitudinal study. Studies in Higher Education, 45(7), 1533-1546.

[8] Sainio, P., Eklund, K., Hirvonen, R., Ahonen, T., & Kiuru, N. (2021). Adolescents' academic emotions and academic achievement across the transition to lower secondary school: The role of learning difficulties. Scandinavian Journal of Educational Research, 65(3), 385-403.

[9] Chen, S. M., Fang, J., Wang, L. M., & Liu, C. L. (2020). Rest a while and run a mile: Relationship between distraction and negative emotions among college students in China. PLoS One, 15(9), e0236030.

[10] Leong, S. C., Tang, Y. M., Lai, C. H., & Lee, C. K. M. (2023). Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing. Computer science review, 48, 100545.

[11] Al-Saadawi, H. F. T., Das, B., & Das, R. (2024). A systematic review of trimodal affective computing approaches: Text, audio, and visual integration in emotion recognition and sentiment analysis. Expert Systems with Applications, 124852.

[12] Taverner, J., Vivancos, E., & Botti, V. (2020). A multidimensional culturally adapted representation of emotions for affective computational simulation and recognition. IEEE Transactions on Affective Computing, 14(1), 761-772.

[13] Järvelä, S., Gašević, D., Seppänen, T., Pechenizkiy, M., & Kirschner, P. A. (2020). Bridging learning sciences, machine learning and affective computing for understanding cognition and affect in collaborative learning. British Journal of Educational Technology, 51(6), 2391-2406.

[14] Cheng Lin, K., Huang, T. C., Hung, J. C., Yen, N. Y., & Ju Chen, S. (2013). Facial emotion recognition towards affective computing - based learning. Library Hi Tech, 31(2), 294-307.

[15] Popescu, R., Ponescu, D., Roibu, H., & Popescu, L. C. (2018, September). Smart classroom-affective computing in present-day classroom. In 2018 28th EAEEIE Annual Conference (EAEEIE) (pp. 1-9). IEEE.

[16] Wang, J. (2025). Emotion Recognition and Regulation in English Teaching Based on Emotion Computing Technology. International Journal of High Speed Electronics and Systems, 34(01), 2540137.

[17] Na, W. (2020). A Data Mining Method for Students' Behavior Understanding. International Journal of Emerging Technologies in Learning, 15(6).

[18] Chuanggao Lin,Fangda Zhang,Yong Zhang & Tingru Zhang. (2025) .Vigilance decrement in automated vs. manual driving: Insights from multimodal data. Transportation Research Part F: Psychology and Behaviour,112,207-220.

[19] Xia Kaiwen,Ren Rui & Liu Feng. (2022) .Numerical analysis of mechanical behavior of stratified rocks containing a single flaw by utilizing the particle flow code. Engineering Analysis with Boundary Elements,137,91-104.

[20] Aijia Xing,Yuchen Sun,Fayi Wei,Lu Yang,Hongwei Zheng & Changhu Xue. (2025) .Deep-learning-assisted chemo-responsive alizarin red S-based hydrogel sensor for the rapid freshness sensing of aquatic product. Food Research International,208,116241-116241.

[21] Soohyun Kim,Young Geun Kim & Yuanjia Wang. (2024) .Temporal generative models for learning heterogeneous group dynamics of ecological momentary assessment data. Biometrics,80(4).

[22] Yangfan Zhou,Jianchun Fan,Baoqian Dai,Shengnan Wu,Rujun Wang,Xinwei Yin... & Xiaofeng Zhang. (2025) .Risk analysis of urban low-pressure natural gas networks based on hybrid dynamic Bayesian networks. Journal of Loss Prevention in the Process Industries,96,105649-105649.