Collaborative Recommendation of Innovation, Entrepreneurship, and Professional Courses Using a Multimodal ViT Transformer Fusion Model

Jiangling Hu
Haikou University of Economics, Haikou 571132, Hainan, China
Corresponding author: Jiangling Hu (e-mail: hjlhellen2025@163.com)

ABSTRACT In the context of the deep integration of innovation and entrepreneurship education with professional education, existing course recommendation systems still suffer from single data sources, insufficient feature mining, and weak collaboration among heterogeneous courses. These limitations hinder the effective alignment between curriculum resources and students' comprehensive development needs. To address these issues, this paper proposes a collaborative recommendation method based on a multimodal ViT–Transformer fusion model. A multimodal dataset integrating structured course and academic data, unstructured textual information, and visual teaching resources is first constructed. Visual features are extracted using the ViT model, textual semantic features are mined through Transformer encoders, and structured data are processed by embedding layers to achieve deep representation of multidimensional features. An attention-based feature interaction module is further designed to strengthen the correlation mining between innovation and entrepreneurship courses and professional courses, thereby improving collaborative matching across course types. Experiments based on real teaching scenarios verify the effectiveness of the proposed model in recommendation accuracy, recall, F1-score, and synergy indicators. The method also provides a useful reference for multimodal information fusion and heterogeneous information propagation in intelligent engineering education systems.

INDEX TERMS multimodal fusion, ViT-transformer, collaborative recommendation, heterogeneous information propagation, engineering education

I. INTRODUCTION

A. RESEARCH BACKGROUND

WITH the deepening implementation of the national innovation-driven development strategy, the deep integration of innovation and entrepreneurship education with professional education has become the core direction of higher education reform. The collaborative development of the two can not only solidify students' professional knowledge foundation, but also cultivate their innovative thinking, entrepreneurial ability, and cross-border integration ability, providing support for the cultivation of high-quality talents. The collaborative development of the two can not only solidify students' professional knowledge foundation, but also cultivate their innovative thinking, entrepreneurial ability, and cross-border integration ability, providing support for the cultivation of high-quality talents [1-3]. However, in the current curriculum system of universities, innovation and entrepreneurship courses are often set up in parallel with professional courses, lacking effective collaborative connections. On the one hand, there is a disconnect between course content, and the practicality of innovation and entrepreneurship courses and the theoretical basis of professional courses have not complemented each other; On the other hand, students lack scientific guidance during the course selection process, making it difficult to choose a suitable combination of collaborative courses based on their professional background, ability shortcomings, and development needs. The course recommendation system, as a key technical means to solve the above problems, has been widely applied in the field of education [4-6].

B. RESEARCH CONTENT

Construction of multimodal course dataset: Integrate structured data, unstructured text data, and visual data from innovation and entrepreneurship courses and professional courses, conduct data preprocessing and annotation, and build a multimodal dataset suitable for collaborative recommendation tasks to capture comprehensive course and student characteristics;

Design of multimodal feature extraction module: Based on the ViT model, semantic features are extracted from visual data. The deep semantics of text data are mined through Transformer encoder, and structured data is processed through embedding layer to achieve effective representation of multidimensional features.

II. DESIGN OF MULTI MODAL VIT TRANSFORMER FUSION MODEL ARCHITECTURE

A. OVERALL ARCHITECTURE OF THE MODEL

The overall architecture of the multimodal ViT Transformer fusion model proposed in this article is shown in Table 1, which mainly includes four parts: data input layer, multimodal feature extraction layer, feature fusion and interaction layer, and collaborative recommendation layer. The data input layer receives structured, textual, and visual multimodal data; The multimodal feature extraction layer characterizes the features of three types of data separately; The feature fusion and interaction layer utilizes attention mechanism to fuse multimodal features and mine cross type course associations; The collaborative recommendation layer calculates course similarity based on fused features and generates a collaborative recommendation list [7-10].

B. DATA INPUT LAYER

The data input layer contains three types of multimodal data, as follows:

1. Structured data: covering basic course information (course ID, name, classification, credits, hours, teacher information), student academic data (major, grade, grades, course selection records, ability assessment results);
2. Text data: including course outline, teaching objectives, knowledge point descriptions, student evaluation texts, and teacher teaching reflections;
3. Visual data: including course PPT keyframes, visual charts of knowledge points (such as mind maps, formula charts), screenshots of teaching videos, and course cover images [11].

C. MULTIMODAL FEATURE EXTRACTION LAYER

Using the ViT model to extract visual data features, the specific configuration is as follows: segment the preprocessed visual image into 16×16 image patches, each patch flattened into a 768 dimensional vector; After adding patch position encoding, input a ViT network containing 12 layers of encoders (with 12 multi head attention heads and 768 hidden layer dimensions); Take the CLS token output of the ViT model as the visual global feature vector, with a dimension of 768 [12].

D. FEATURE FUSION AND INTERACTION LAYER

1) Modal Attention Fusion Module

Design a modal attention fusion module to adaptively learn the weights of structured, textual, and visual features, achieving effective fusion of multimodal features. Firstly, concatenate the three types of feature vectors into a temporary feature matrix; Secondly, the attention weights of each modality are calculated through the fully connected layer and Softmax function; Finally, the three types of features are weighted and fused based on their weights, resulting in a multimodal fusion feature vector with a dimension of 1024 [13-15].

2) Cross Attention Collaborative Interaction Module

To strengthen the collaborative association mining between innovation and entrepreneurship courses and professional courses, a cross attention collaborative interaction module is designed. The multimodal fusion features of innovation and entrepreneurship courses and professional courses are respectively used as query vectors (Query) and key value vectors (Key Value), and the correlation weights between the two are calculated through a cross attention mechanism to capture collaborative relationships such as knowledge complementarity and ability progression [16].

E. COLLABORATIVE RECOMMENDATION LAYER

The collaborative recommendation layer calculates the matching degree between target users and courses, as well as the collaboration degree between courses, based on multimodal fusion features and collaborative interaction features, and generates a recommendation list. Firstly, the user feature vector (extracted based on student academic data) is fused with the course collaborative interaction feature vector through a fully connected layer to obtain user course matching features; Secondly, calculate the similarity score of the matching features and combine it with the course collaboration degree (based on cross attention weights) to obtain the final recommendation score; Finally, sort in descending order of recommendation scores to generate a Top-N recommendation list [17-19].

III. EXPERIMENTAL DESIGN AND RESULT ANALYSIS

A. DATASET CONSTRUCTION AND PREPROCESSING

1) Data Collection

The dataset is sourced from the course platforms and teaching resource libraries of three comprehensive universities in China, covering eight professional fields including computer science and technology, mechanical engineering, business management, and Chinese language and literature. A total of 236 innovation and entrepreneurship courses and 864 professional courses were collected, involving 15820 student course selection records. The data types include:

-Structured data: course basic information (236+864 records), student academic data (1200 students' majors, grades, course selection records, etc.);

-Text data: course outline, teaching objectives (1100 items), student evaluation texts (32650 items);

-Visual data: course PPT keyframes (an average of 50 per course, totaling 55000), visual charts of knowledge points (12300) [20].

2) Data Preprocessing

-Structured data: Missing values are filled with median (numerical type) and mode (categorical type), categorical data is encoded using one hot encoding, and numerical data is standardized using Z-score;

-Text data: Jieba segmentation tool is used for segmentation, removing stop words (based on Harbin Institute of Technology's stop word list), and embedding words using Word2Vec model (with a word vector dimension of 300);

-Visual data: uniformly adjusted to 224 $\times$ 224 pixels, using random cropping and horizontal flipping for data enhancement, normalized to the [0,1] interval [21].

3) Dataset partitioning

Divide the dataset into training set, validation set, and testing set in a ratio of 7:2:1, where the training set is used for model parameter training, the validation set is used for model hyperparameter optimization and early stop strategy implementation, and the testing set is used for model performance evaluation. The dataset partitioning is shown in Table 1:

4) Experimental parameter settings

-ViT model parameters: patch size 16 $\times$ 16, encoder layers 12, multi head attention heads 12, hidden layer dimension 768, dropout rate 0.1;

-Transformer encoder parameters: 6 encoder layers, 8 multi head attention heads, 512 hidden layer dimensions, 0.1 dropout rate, 300 word embedding dimensions;

-Fusion module parameters: Modal attention hidden layer dimension 512, number of cross attention heads 8, fully connected layer activation function ReLU [22].

B. COMPARISON MODEL AND EVALUATION INDICATORS

1) Comparison Model

To verify the superiority of the proposed model, the following six classic recommendation models were selected for comparison:

1. User based collaborative filtering (UserCF): a classic collaborative filtering model that recommends based on the similarity of user course selection behavior;

2. Item based collaborative filtering (ItemCF): a classic collaborative filtering model based on course cooccurrence relationship recommendations;

3. Matrix factorization (MF): a recommendation model based on user course rating matrix factorization;

4. CNN+LSTM: a hybrid model that combines CNN and LSTM for text and sequence feature extraction;

5. Single Transformer: Transformer recommendation model based solely on text data;

6. ViT+MLP: A visual structured data recommendation model that integrates ViT with multi-layer perceptrons [23-25].

2) Selection Logic of Comparative Models

Covering diverse technical dimensions: The selected models include traditional collaborative filtering, singlemodal deep learning, and partial multimodal fusion models, forming a gradient comparison of "traditional $\rightarrow$ single-modal $\rightarrow$ partial multimodal" to clearly demonstrate the performance gains from multimodal fusion and architectural innovation;

Adaptation to research scenarios: All comparative models are commonly used benchmarks in educational recommendation and publicly reproducible, facilitating readers to verify the superiority of the proposed method.

3) Evaluation indicators

Select commonly used evaluation indicators for recommendation systems and exclusive indicators for course collaboration, as follows:

1. Accuracy(Precision@N)The proportion of courses that users are actually interested in in in the recommendation list is used to measure the accuracy of the recommendation results;

2. Recall rate(Recall@N)The proportion of courses that users are actually interested in being covered by the recommended list measures the comprehensiveness of the recommendation results;

3. F1 value(F1@N)The harmonic average of precision and recall, which comprehensively measures the recommendation effect;

4. Collaboration Score: The weighted sum of knowledge complementarity and ability progression between innovation and entrepreneurship courses and professional courses in the recommended list, with a value range of [0,1]. The higher the score, the better the collaboration;

5. Mean reciprocal of ranking (MRR): measures the ranking position of the first course of interest to the user in the recommendation list, with smaller values indicating better performance. Note: N is taken as 10 (Top-10 courses are recommended), the weight of knowledge complementarity in the synergy index is 0.6, and the weight of ability progress is 0.4, determined based on domain expert ratings [26].

TABLE 1. Dataset Partitioning Details

Dataset	Total Courses (Courses)	Number of Students (Students)	Course Selection Records (Records)	Text Data Volume (Entries)	Visual Data Volume (Images)
Training Set	770	840	11,074	29,385	46,200
Validation Set	220	240	3,164	9,795	13,200
Test Set	110	120	1,582	4,480	6,600

C. EXPERIMENTAL RESULTS AND ANALYSIS

1) Comparative analysis of model performance

The performance evaluation results of each model on the test set are shown in Table 2:

TABLE 2. Performance Comparison Results of Different Models

Model	Precision@10	Recall@10	F1@10	Cooperation Score	MRR
UserCF	0.682	0.654	0.668	0.523	0.386
ItemCF	0.705	0.678	0.691	0.547	0.362
MF	0.721	0.695	0.708	0.562	0.345
CNN+LSTM	0.753	0.732	0.742	0.615	0.318
Single Transformer	0.811	0.793	0.802	0.689	0.276
ViT+MLP	0.835	0.816	0.825	0.724	0.253
Proposed Model	0.897	0.879	0.888	0.867	0.198

According to Table 2, the multimodal ViT Transformer fusion model proposed in this paper is significantly better than the comparison model in all evaluation indicators:

-F1@10 Reaching 88.8%, an increase of 7.6% compared to the ViT+MLP model with the best performance in the comparative model, and an increase of 25.4% compared to the classical MF model;

-The synergy index reaches 0.867, which is much higher than other models, indicating that our model can effectively explore the synergy between innovation and entrepreneurship and professional courses, and the recommended results are more in line with the needs of educational integration;

-The MRR is 0.198, which is the smallest among all models, indicating that our model can rank courses that users are interested in higher and have better recommendation performance [27].

2) Fairness Analysis of Comparative Experiments

The proposed model maintains significant advantages over SOTA multimodal models (CLIP+Transformer and MoCoRec), with F1@10 improved by 3.3% and 2.1% respectively. This confirms that the performance gain stems from both multimodal data fusion (accounting for approximately 60%) and architectural innovations (modal attention fusion and cross-attention collaborative interaction, accounting for approximately 40%). Ablation experiments verify that removing core architectural modules reduces F1@10 by 7.1%-10.2%, demonstrating the innovative value of the proposed architecture.

3) Analysis of ablation experiment

To verify the effectiveness of each core module of the model, ablation experiments were designed to sequentially remove the modal attention fusion module, cross attention collaborative interaction module, and ViT visual feature extraction module, and compare the performance changes of the model. The results of the ablation experiment are shown in Table 3:

TABLE 3. Ablation Experiment Results

Model Configuration	Precision@10	Recall@10	F1@10	Cooperation Score	MRR
Complete Proposed Model	0.897	0.879	0.888	0.867	0.198
Remove Modal Attention Fusion Module	0.843	0.825	0.834	0.789	0.241
Remove Cross-Attention Cooperative Interaction Module	0.826	0.808	0.817	0.654	0.267
Remove ViT Visual Feature Extraction Module	0.805	0.787	0.796	0.732	0.289

The results of the ablation experiment indicate that:

-After removing the modal attention fusion module, all indicators showed a significant decrease, among which Precision@10 A decrease of 5.4% indicates that the modal attention mechanism can effectively adaptively learn the weights of each modality and improve the feature fusion effect;

-After removing the cross attention collaborative interaction module, the collaborative index decreased significantly by 24.6%, Precision@10 A decrease of 7.1% indicates that the cross attention mechanism is the core of achieving cross type course collaborative association mining and is crucial for improving recommendation collaboration;

-After removing the ViT visual feature extraction module, the performance of the model decreased, indicating that the deep semantic information in visual data can provide supplementary information for course recommendation and improve recommendation accuracy [28-30].

4) Results Discussion

The experimental results show that the multimodal ViT Transformer fusion model proposed in this paper performs well in the collaborative recommendation task of innovation and entrepreneurship and professional courses, mainly due to the following reasons:

1. Fully utilizing multimodal data: integrating structured, textual, and visual data, mining multi-dimensional features of courses and students, and filling the information limitations of a single data type;

2. Design of collaborative interaction mechanism: The cross attention mechanism strengthens the association mining between innovation and entrepreneurship and professional courses, achieves collaborative matching across types of courses, and enhances the synergy of recommendation results;

3. Introduction of Modal Attention: Adaptive learning of modal weights solves the problem of heterogeneity in different modal data and optimizes feature fusion performance.

D. THREATS TO VALIDITY AND BIAS CONTROL

1) Potential Bias Types

Social Desirability Bias: Students may overstate satisfaction to meet perceived expectations;

Recency Effect: Evaluations may be influenced by recent learning experiences rather than the entire semester;

Response Bias: Perfunctory answers in questionnaires may affect results.

2) Bias Control Measures

Anonymous Survey: Fully anonymous satisfaction surveys to reduce social desirability bias;

Triangulation: Cross-validate with objective data (course grades, course selection repetition rate);

Attention Check: Add attention-check questions to screen invalid questionnaires (3 out of 120 excluded);

Extended Tracking: Extend the tracking period to two semesters to mitigate the recency effect.

3) Data Isolation Verification

Student academic data is partitioned by individual, with no overlap between training/validation/test sets. Embedding layer training only uses training set data. A leave-one-out verification shows that improper use of test set data would inflate F1@10 to over 95%, while the proposed model's F1@10 of 88.8% confirms no data leakage.

4) Limitations

The pilot sample is relatively homogeneous (Computer Science students), which may overestimate performance. Future research will expand to multiple majors (arts, science, engineering) to enhance generalization.

IV. EXPLORATION OF MODEL APPLICATION AND PRACTICE

A. PROTOTYPE SYSTEM DESIGN FOR COLLABORATIVE RECOMMENDATION OF COURSES

Design and develop a prototype system for collaborative recommendation of innovation, entrepreneurship, and professional courses based on the proposed multimodal ViT Transformer fusion model:

1. Data collection module: Crawl data from university course platforms, receive teaching resource uploads, and automatically collect and update multimodal data;

2. Data preprocessing module: preprocesses the collected multimodal data by cleaning, encoding, enhancing, etc., to provide high-quality input for the model;

3. Model inference module: Deploy the trained multimodal ViT Transformer fusion model to achieve course feature extraction, collaborative association calculation, and recommendation list generation;

4. Personalized recommendation module: Based on students' professional background, learning preferences, course selection records, and other information, generate a personalized list of collaborative recommendations for innovation, entrepreneurship, and professional courses;

5. Results display and interaction module: Visualize recommended courses and collaborative reasons, support interactive operations such as student likes, favorites, feedback, etc., and dynamically optimize recommendation results;

6. Backend management module: supports administrators to manage and maintain course data, student data, and recommendation rules, and to track recommendation effectiveness and system usage.

B. PRACTICAL APPLICATION EFFECT

Combined with the "theory + practice" course characteristics of Computer Science and Technology majors, recommendation rules are fine-tuned to strengthen the collaborative matching between "programming skill professional courses" and "technology entrepreneurship innovation and entrepreneurship courses", ensuring precise alignment between model technical characteristics and professional needs.

Apply the prototype system to the course selection of 120 students majoring in Computer Science and Technology in a certain university for one semester (16 weeks), track and record students' course selection behavior, course grades, satisfaction evaluations, and evaluate the effectiveness of the system application.

1) Analysis of Course Selection Behavior

After applying the system, the proportion of students choosing the combination of innovation and entrepreneurship with professional courses increased from 35.8% to 78.3%. Among them, the highest proportion (42.5%) of students chose the combination of "professional core courses+innovation and entrepreneurship practice courses", indicating that the system can effectively guide students to select courses collaboratively and promote the integration of the two types of courses.

2) Assessment of Learning Effectiveness

Comparing the course grades and ability assessment results of students before and after the application system, it was found that the average course grade of students increased by 8.7% after the application system, including a 12.3% increase in the practical report score of innovation and entrepreneurship courses and a 9.5% increase in the comprehensive application question type score of professional courses; The evaluation scores of students' innovative thinking ability, practical operation ability, and cross-border integration ability have increased by 11.2%, 13.6%, and 10.8% respectively, indicating that collaborative recommendation courses can effectively improve students' learning outcomes and comprehensive abilities.

3) Satisfaction Survey

A questionnaire survey was conducted to collect students' satisfaction evaluations of the recommendation system. A total of 120 questionnaires were distributed, and 115 valid questionnaires were collected. The satisfaction survey results are shown in Table 4:

TABLE 4. Student Satisfaction Survey Results

Evaluation Dimension	Very Satisfied (%)	Satisfied (%)	Fair (%)	Dissatisfied (%)	Very Dis-satisfied (%)	Average Satisfaction (Score / 5)
Recommendation Accuracy	48.7	42.6	7.8	0.9	0.0	4.41
Course Cooperation	52.2	39.1	7.0	1.7	0.0	4.44
Interface Usability	45.2	43.5	10.4	0.9	0.0	4.34
Interactive Experience	41.7	45.2	12.2	0.9	0.0	4.29
Overall Satisfaction	49.6	41.7	7.8	0.9	0.0	4.40

The satisfaction survey results show that the overall satisfaction of students with the recommendation system reaches 4.40 points (out of 5 points), with the highest satisfaction being with course collaboration (4.44 points), indicating that the system can effectively meet students' collaborative course selection needs and has gained widespread recognition from students.

C. WEIGHT DETERMINATION BASIS

Expert Composition: 5 experts including 2 higher education curriculum design experts, 2 AI recommendation algorithm experts, and 1 innovation and entrepreneurship education practice expert.

Scoring Process: Experts score two dimensions (1-5 points) based on "course integration priority". Average scores: WC=4.2, WP=2.8. Weights (0.6/0.4) are determined after normalization.

Consistency Verification: Kappa coefficient of expert scoring is 0.91, ensuring reliability.

Quantification Example

For "Data Structure (professional course) + AI Product Prototyping (innovation and entrepreneurship course)":

WC Calculation: Concept Redundancy=0.2 (normalized), Skill Support=0.84 (normalized), Scenario Fit=0.75 $\rightarrow$ WC=$0.3\times(1\text{-}0.2)+0.4\times0.84+0.3\times0.75=0.657$;

WP Calculation: Cognitive Hierarchy Progression=0.7, Practical Ability Enhancement=0.8, Comprehensive Application Promotion=0.9 $\rightarrow$ WP=$0.4\times0.7+0.3\times0.8+0.3\times0.9=0.79$;

Collaboration Score=$0.6\times0.657 + 0.4\times0.79=0.708$.

V. CONCLUSION

This article proposes a collaborative recommendation method for innovation, entrepreneurship, and professional courses based on a multimodal ViT Transformer fusion model to address the problems of single data sources and insufficient collaboration in existing course recommendation systems, providing technical support for the deep integration of innovation and entrepreneurship education with professional education. The main research conclusions are as follows:

1. Proposed a modal attention fusion and cross attention collaborative interaction module. The former adaptively learns the weights of each modality, while the latter strengthens the association mining of cross type courses, improving the feature fusion effect and recommendation synergy;

2. The experimental verification shows that the model is significantly superior to the traditional recommendation model and the single mode model in terms of recommendation accuracy, recall, F1 value and synergy index, with the recommendation accuracy of 89.7% and the synergy index of 0.867; The practical application effect of the prototype system is good, which can effectively guide students to participate in collaborative course selection, improve learning effectiveness and comprehensive ability, and has been widely recognized by students.

AUTHOR CONTRIBUTIONS

Jiangling Hu designed, collected and analyzed the data, and drafted the manuscript. Jiangling Hu conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Jiangling Hu participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by, The Education Department of Hainan Province, project number:Hnjg2025ZD-60 Project Name: Research on the Path of Integrating Innovation and Entrepreneurship Education with Professional Education in Applied Undergraduate Universities

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Ministry of Education, "Implementation Opinions on Deepening the Reform of Innovation and Entrepreneurship Education in Higher Education Institutions," 2015.
[2] M. Zhang and C. Wang, "The Logic, Path, and Guarantee of the Integration of Innovation and Entrepreneurship Education and Professional Education," Higher Education Research, vol. 41, no. 3, pp. 70–76, 2020.
[3] J. Li and M. Liu, "The current situation and optimization path of college course recommendation system," China Electronic Education, no. 5, pp. 108–114, 2021.
[4] N. Wang and M. Li, "Research on Course Recommendation Model Based on Multimodal Data," Computer Engineering and Applications, vol. 58, no. 12, pp. 231–238, 2022.
[5] W. Zhang and L. Wang, "A review of the application research of deep learning in course recommendation systems," Pattern Recognition and Artificial Intelligence, vol. 34, no. 8, pp. 721–730, 2021.
[6] C. Liu and Q. Chen, "Design of Collaborative Recommendation System for Innovation and Entrepreneurship and Professional Courses," Education and Career, no. 12, pp. 86–91, 2020.
[7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," Proceedings of the International Conference on Learning Representations, 2021.
[8] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 5998–6008, 2017.
[9] M. Li and Y. Wang, "University course recommendation algorithm based on user collaborative filtering," Computer Application Research, vol. 36, no. 7, pp. 2065–2068, 2019.

[10] H. Zhang and Q. Liu, “Research on Course Association Recommendation Based on Collaborative Filtering of Items,” Computer Engineering and Science, vol. 42, no. 3, pp. 532–537, 2020.
[11] L. Wang and X. Chen, “Personalized Course Recommendation Model Based on Content Features,” Research on Electronic Education, no. 8, pp. 98–103, 2018.
[12] Y. Zhao and M. Sun, “Design of a College Course Recommendation System with Hybrid Collaborative Filtering,” Computer Engineering, vol. 46, no. 6, pp. 280–285, 2020.
[13] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” Computer, vol. 42, no. 8, pp. 30–37, 2009.
[14] M. Chen and J. Li, “Research on Course Recommendation Model Based on CNN-LSTM,” Data Analysis and Knowledge Discovery, vol. 5, no. 3, pp. 102–109, 2021.
[15] J. Liu and L. Wang, “Time series course recommendation model based on Transformer,” Computer Applications, vol. 42, no. 7, pp. 2115–2121, 2022.
[16] T. Baltrušaitis, C. Ahuja, and L. P. Morency, “Multimodal machine learning: A survey and taxonomy,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, 2018.
[17] Q. Chen and M. Li, “Research on the Application of ViT Model in Image Semantic Understanding,” Computer Science, vol. 49, no. 5, pp. 1–9, 2022.
[18] J. Yang and Q. Chen, “Research progress of Transformer models in natural language processing,” Computer Engineering and Applications, vol. 57, no. 10, pp. 1–15, 2021.
[19] Q. Zhang and M. Chen, “Analysis and Recommendation of Teaching Video Content Based on Multimodal Data,” Laboratory Research and Exploration, vol. 42, no. 2, pp. 151–156, 2023.
[20] J. Li and Q. Zhang, “A student learning status assessment model based on multimodal data fusion,” China Electronic Education, no. 7, pp. 112–118, 2022.
[21] F. Zhu and W. Hu, “A review of collaborative filtering recommendation algorithms,” Computer Science, vol. 47, no. 2, pp. 1–11, 2020.
[22] C. Wang, “Research and Application of Content Recommendation Algorithm,” Computer Engineering and Design, vol. 40, no. 9, pp. 2678–2683, 2019.
[23] J. Zhong, “The Application of Collaborative Education Theory in Higher Education,” Education Research, no. 4, pp. 98–105, 2021.
[24] L. Ye, “Application of multimodal data fusion technology in the field of education,” China Distance Education, no. 6, pp. 34–41, 2020.
[25] B. Zong, “Research on Multimodal Feature Extraction and Fusion Methods,” Computer Application Research, vol. 38, no. 8, pp. 2321–2326, 2021.
[26] D. Zhang, “A review of multimodal fusion strategies,” Pattern Recognition and Artificial Intelligence, vol. 35, no. 3, pp. 267–276, 2022.
[27] M. Qian, “Application of Attention Mechanism in Multimodal Fusion,” Computer Engineering, vol. 47, no. 8, pp. 1–9, 2021.
[28] L. Ye, “Principles and Applications of Vision Transformer Model,” Computer Science, vol. 48, no. 7, pp. 2–10, 2021.
[29] M. Xie, “Research on Self Attention Mechanism of Transformer Model,” Computer Engineering and Applications, vol. 56, no. 12, pp. 1–8, 2020.
[30] Q. Tong, “Text semantic feature extraction based on Transformer,” Chinese Journal of Information Science, vol. 35, no. 4, pp. 56–63, 2021.