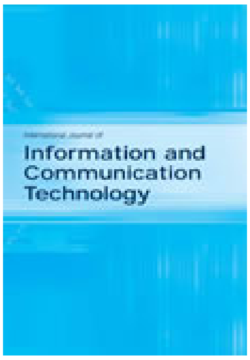
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Computer learning career path optimisation utilising multi-modal large models and privacy-preserving collaborative computing

Xuesong Yang

DOI: 10.1504/IJICT.2025.10072945

Article History:

Received:	22 June 2025
Last revised:	14 July 2025
Accepted:	15 July 2025
Published online:	08 September 2025

Computer learning career path optimisation utilising multi-modal large models and privacy-preserving collaborative computing

Xuesong Yang

College of Electrical Engineering,
Northwest Minzu University,
Gansu 730000, China
Email: xuesongyang2025@163.com

Abstract: As computer technology advances, there is a growing need for personalised learning path planning for learners. Traditional methods fall short in accuracy and adaptability. This study introduces MPCO, a computer course learning path optimisation model powered by a multi-modal large model and privacy computation. The multi-modal large model integrates text, images, and other info to better understand learners' knowledge levels and cognitive preferences. Privacy computation technology ensures the safe storage and compliant sharing of learning data, reducing the risk of data privacy breaches. Experiments show that this method achieves higher accuracy, adaptability, and data security in learning path optimisation tasks through the collaborative driving of multimodal large models and privacy computing, effectively improving the planning effect of computer course learning paths.

Keywords: multi-modal large models; privacy computing; computer courses; learning path optimisation; personalised learning; Java.

Reference to this paper should be made as follows: Yang, X. (2025) 'Computer learning career path optimisation utilising multi-modal large models and privacy-preserving collaborative computing', *Int. J. Information and Communication Technology*, Vol. 26, No. 32, pp.83–100.

Biographical notes: Xuesong Yang received his Master's degree from Universiti Utara Malaysia in 2010. He is currently a Lecturer in the College of Electrical Engineering at Northwest Minzu University. His research interests include information systems and information security.

1 Introduction

Nowadays, computers are super important, and people want to learn more about them. But the old ways of planning learning paths are not that great. They cannot really meet the different needs of each learner. Personalised learning path planning is now really important because it can help learners study more efficiently (Zhao et al., 2024). It is all about making the best learning path for each person based on things like what they already know, how they like to learn and how smart they are. But the traditional ways have a lot of problems. They are not very accurate and cannot change much. They often use very little info about the user and some simple rules to plan the path. So the recommendations are not that good. Also, they cannot really adapt to the different needs that come up during learning. They are not flexible or dynamic, with more educational data being shared, data privacy is a big concern. How to keep learner data secure and share it properly without leaks or misuse is a problem that needs to be solved right away. On the bright side, multi-modal models are really good at putting together different types of info. They can handle text, images, videos and so on. By looking at things like learners' written work, code exercises and graphic designs in the course, the model can better understand their learning levels, interests and weak points. This makes it better at planning learning paths and giving more personalised and accurate learning plans. Also, multi-modal large models can create more complete learner profiles by combining different data sources, which is great for personalised learning path planning. Privacy computing helps keep data sharing secure. In personalised learning path planning, a large amount of learner data is involved, including personal information, learning behaviour, knowledge mastery, etc. These data have high privacy sensitivity. Privacy computing technology can achieve secure sharing and effective utilisation of data while protecting learners' privacy, ensuring the security and compliance of learners' data in the learning path planning process. Through homomorphic encryption technology, analysis and computation can be directly performed on encrypted data without the need for decryption, thus protecting the privacy of learners; differential privacy technology ensures the privacy of individual data is not compromised by adding appropriate noise during the data publishing and analysis process; zero knowledge proof technology can prove the authenticity of a statement without leaking any additional information, providing strong support for secure data sharing and identity authentication.

In online education, piecing together scattered learning materials via learning roadmaps to steer learners is crucial for swiftly aligning learners' objectives with suitable content. Due to varying learning capacities, prior knowledge, learning aims, and cognitive tiers among learners, the prevailing rigid and uniform approach to resource selection and learning trajectories often leaves learners struggling to promptly locate appropriate resources (Wang, 2025). Consequently, devising precise personalised learning pathways that respect individual distinctions has emerged as a pivotal focus in personalised learning research. Typically, personalised learning path design approaches fall into three main groups, each enhancing the learning process from unique angles. The learner-trait-centred method, for one, incorporates elements like learning ambitions and hobby inclinations. Scholars build a learner profile by gathering data on learning aims, styles, hobbies, via assessments or surveys. For example, Dwivedi et al. (2018) proposed a learning path recommendation method based on an improved variable length genetic algorithm, aiming to enhance learning effectiveness by suggesting suitable paths. Niknam and Thulasiraman (2020) developed a biologically inspired intelligent learning path recommendation system (LPR) rooted in meaningful learning theory, recommending paths that align with learners' cognitive traits. Vanitha et al. (2019) highlighted the importance of learning goals and knowledge level in path planning, employing a collaborative optimisation algorithm that combines ant colony and genetic algorithms. Nabizadeh et al. (2020) utilised depth-first search algorithm with learning objectives and knowledge graphs to identify numerous course sequences. Nevertheless, creating personalised paths based solely on learner characteristics may neglect knowledge relevance, causing learners to struggle with grasping the entire knowledge structure. Path planning leveraging log data makes use of learners' historical behaviour to understand their traits and recommend needed learning objects. In online learning, historical behavioural data can be fundamental parameters for path construction. For instance, Jiang et al. (2022) explored data-driven personalised path planning based on cognitive diagnostic assessment in MOOCs to offer more precise paths. Liu and Li (2020) adopted a learning path combination recommendation method based on learner log data to suggest paths using historical behaviour. Zhou et al. (2018) applied clustering algorithms and long LSTM recurrent neural networks for path recommendation. Yet, paths derived from group data may lack suitability for all learners. In addition, in the case of insufficient data, there will be a ‘cold start’ problem. Personalised learning path planning based on knowledge graph considers course resources and learner behaviour data, usually using self-designed topology sorting algorithms and optimisation algorithms to serialise course resources based on the relationships and attributes between courses, and planning learning paths from a knowledge structure perspective to ensure learning efficiency (Huang et al., 2014). For example, Zhu et al. (2018) proposed a multi-constraint learning path recommendation algorithm based on knowledge graph. It can tackle the issue where learners struggle to pick suitable learning materials. Shmelev et al. (2015) combined genetic methods and knowledge graph technology to arrange course resources in order as learning paths. However, most knowledge graph-based path generation methods ignore the learners’ changing knowledge states over time, and the learning paths generated at the initial stage of the course may not be suitable for the learners’ entire learning process. In summary, although these methods have made some progress in personalised learning path planning, there are still some problems. Such as coarse-grained planning sequences and limited user data features. Future research should comprehensively consider various information such as learner characteristics, log data, and knowledge graphs to achieve more accurate and comprehensive personalised learning path planning.

In this context, this article proposes a computer course learning path optimisation strategy based on the collaborative drive of multi-modal large models and privacy computing. This strategy aims to fully utilise the powerful information processing capabilities of multi-modal large models and the data security guarantee capabilities of privacy computing, providing learners with more accurate, secure, and personalised learning path planning solutions. Specifically, multi-modal large models integrate learners’ multi-source data, analyse their learning characteristics and needs in depth, and generate personalised learning paths. Meanwhile, privacy computing technology runs through the entire process, ensuring the security and privacy of learners’ data, effectively mitigating the risk of data privacy leakage in traditional methods. The specific trio of homomorphic encryption, differential privacy and zero-knowledge proof was adopted because it uniquely supports encrypted computation on ciphertext, injects calibrated noise for anonymity, and enables verifiable model updates without ever revealing raw learner records – capabilities not jointly offered by alternative techniques such as secure multi-party computation alone.

The main innovations and contributions of this work include:

1 This article innovatively introduces a multi-modal large model to comprehensively process multi-source data such as text assignments, code practices, and graphic design. Unlike traditional single text or few feature analysis methods, this model can comprehensively explore multidimensional features such as learners’ learning level, interest points, and knowledge weaknesses. By deeply integrating multi-modal data and constructing accurate learner profiles, comprehensive basis is provided for

personalised learning path planning, significantly improving the level of personalised planning.

2 This article proposes a data sharing mechanism based on privacy computing to address data privacy issues. By utilising techniques such as homomorphic encryption and differential privacy, secure sharing and effective utilisation of data can be achieved under encrypted conditions. Learner data is always encrypted and can be analysed and calculated without decryption, ensuring the security and compliance of data in learning path planning, eliminating data silos, and ensuring data availability.

3 This article proposes a collaborative optimisation framework for multi-modal large models and privacy computing. The multi-modal large model is responsible for learner feature mining and learning path generation, while privacy computing ensures data sharing and model training security. The two work together to balance the accuracy of personalised recommendations and data privacy protection, achieve complementary advantages, jointly improve the effectiveness of learning path planning, provide high-quality personalised learning solutions for learners, and promote the innovative development of computer education.

2 Relevant technologies

2.1 Transformer

Transformer, a neural network architecture built on the self-attention mechanism, first gained remarkable success in natural language processing (NLP), particularly in machine translation, and has since become widely used across various NLP tasks (Tetko et al., 2020). Currently, transformer is applied in the fields of image recognition and multi-modal emotion recognition, with very good performance. Unlike traditional RNNs and CNNs, transformer networks do not have cyclic and convolutional structures, allowing for parallel processing of each element in a sequence, thereby speeding up training and inference. The transformer model comprises an encoder-decoder architecture (Han et al., 2022). The decoder then uses these vectors and prior predictions to generate the target sequence step by step.

Position encoding plays a vital role in transformers. It injects positional details into each component of the input sequence, enabling the model to recognise sequence order. This can be achieved by incorporating particular combinations of sine and cosine functions into the input vector:

$$PE(pos, 2i) = \sin\left(pos/10{,}000^{2i/d_{model}}\right) \tag{1}$$

$$PE(pos, 2i+1) = \cos\left(pos/10{,}000^{2i/d_{model}}\right) \tag{2}$$

in the context of position encoding, *pos* denotes an element's position, while *i* indicates dimensional info. Per the cosine formula, a linear function can express the encoding of position $i + k$ based on position's encoding.

The self-attention mechanism identifies the relationships between various positions in an input sequence. Within each encoder and decoder transformer block, the multi-head self-attention mechanism transforms the input sequence into key, query, and value triplets. Calculating the dot product of queries and keys produces attention score vectors. Multiplying these by the corresponding value vectors generates the self-attention vectors. The multi-head self-attention mechanism divides the attention calculation across multiple heads, computes distinct attention score vectors for each head, combines the resulting self-attention vectors, and applies a fully connected layer for dimensionality reduction. Specific calculations are as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{4}$$

$$MultiHead(Q, K, V) = Concat\left(head_1, head_2, \ldots, head_n\right)W^O \tag{5}$$

where W_i^Q, W_i^K, W_i^V and W^O represent different weight matrices, and $head_i$ represents the score vector calculated by the i^{th} head attention module. The multi-head attention mechanism maps the original vector collection to several subspaces and independently computes attention weights in each subspace. It linearly transforms the input *query*, *key*, and *value* vectors into multiple low dimensional spaces, and then separately calculates the dot product attention weights in each space. Finally, the attention information obtained from each subspace is concatenated and subjected to a linear transformation to obtain the final output vector.

Feedforward networks are essential components of the transformer architecture. Within each block, these networks take the output from the multi-head self-attention mechanism as their input. They then employ two linear transformations in conjunction with a nonlinear activation function to project the input into a new representation vector. The specific formulation is as follows:

$$FFN(x) = \max\left(0, xW_1 + b_1\right)W_2 + b_2 \tag{6}$$

where W_1 and W_2 represent two different linear functions, and b_1 and b_2 are two bias constants.

Residual connections allow information to flow directly from one layer to the next by adding the input to the output. The specific formula is as follows:

$$Outlayer = LaterNorm(x + Sublayer(x)) \tag{7}$$

among them, *LaterNorm*() is the layer normalisation function, and *Sublayer*(x) represents the function implemented by each sublayer itself.

2.2 Multi-modal fusion technology

Human beings live in a multi-modal environment, where cognition and behaviour exhibit multi-modality. Recently, the development of sensor technology has led to an increase in data quality, quantity, and variety. For humans, different types of data collected by multiple sensors can provide diverse perspectives on specific objects or scenes, building more comprehensive cognition (Zhang et al., 2023). Inspired by this, researchers in the field of artificial intelligence have conducted research on multi-modal technology, which involves constructing artificial intelligence models that can understand multiple modal data to enhance robustness. The multi-modal model aims to map different modal data to the same space, learn multi-modal collaborative knowledge through interaction and alignment between modalities, and explore potential correlation information between modalities (Goyal et al., 2022). Current studies mainly concentrate on visual, textual, and audio modalities. They have attained top notch outcomes in various comprehension and generation tasks like text search, visual Q&A, and video captioning. It is considered as an exploration of the path from expert systems to general artificial intelligence.

Recently, attention mechanism has become one of the most important technologies in multi-modal fusion. The attention mechanism can help the model make more accurate predictions without significantly increasing computational resource consumption (Tan et al., 2021). Many different forms of attention mechanisms have been proposed, such as channel and spatial attention mechanisms, self-attention mechanisms, and cross modal attention mechanisms. In convolutional neural networks, by applying weights to different channels and spatial positions of features, the model can learn two important questions: ‘what to look at’ and ‘where to look at’. In other words, different channels of convolutional neural network features usually correspond to different feature maps, which can capture different semantic information. By using channel and spatial attention mechanisms, meaningful features can be highlighted from both channel and spatial dimensions, and invalid information can be filtered out. Given a feature map $f \in \mathbb{R}^{C \times H \times W}$ with C channels and a size of $H \times W$, the convolutional attention module first infers the channel attention map $A_c \in \mathbb{R}^{C \times 1 \times 1}$ and spatial attention map $A_s \in \mathbb{R}^{1 \times H \times W}$ (Liu et al., 2023). The overall feature optimisation process based on attention mechanism can be expressed as follows:

$$f' = A_c(f) \otimes f \tag{8}$$

$$f'' = A_s(f') \otimes f' \tag{9}$$

here $\otimes$ represents element by element multiplication, and f'' is the final optimised output feature.

Computing the attention map integrates average pooling and max pooling features. These pooled features are then forwarded to a common multi-layer perceptron (MLP) network to generate the final channel attention map. Here is how the channel attention map is computed:

$$A_c(f) = \sigma(MLP(AvgPool(f)) + MLP(MaxPool(f))) \tag{10}$$

where σ denotes the Sigmoid function. To compute the spatial attention map, the input feature map undergoes average pooling and max pooling along the channel dimension to generate two feature maps, $f_{avg}^s \in \mathbb{R}^{1 \times H \times W}$ and $f_{\max}^s \in \mathbb{R}^{1 \times H \times W}$.

$$A_s(f) = \sigma\left(Conv\left[f_{avg}^s; f_{\max}^s\right]\right) \tag{11}$$

The core idea of multi-modal attention mechanism is based on its powerful non-local attention ability, which utilises information from another modality to enhance the features of the main task. This unique approach can effectively explore the potential correlations between different modalities, utilising the effective information of different modalities to supplement and enhance the features of the main modalities. Specifically, the chosen triad of textual assignments, code snapshots and graphic artefacts jointly captures declarative, procedural and visual dimensions of Java mastery – intersecting these signals uncovers latent misconceptions that any single modality would miss, thus providing the most comprehensive basis for sequencing learning paths. This method is mainly applied in transformer-based models, using multi-modal attention mechanisms to mine supplementary information from other modalities. The feature enhancement method of multi-modal attention mechanism has multiple advantages. This approach greatly boosts the feature expression of multi-modal data, allowing the model to more effectively comprehend and combine cross modal information. It also helps break through single-modal data limits and fully unlock the potential of multi-modal data. Furthermore, by dynamically adjusting the attention weights between different modalities, it can adapt to different task requirements and data characteristics, and has high flexibility and robustness.

2.3 Privacy computing

Privacy computing is a series of information technologies aimed at analysing and calculating data without leaking raw data, in order to achieve 'usable but invisible' data in the process of circulation and fusion (Li et al., 2019). In terms of technical principles, privacy computing technology is cross integrated with many disciplines such as artificial intelligence, cryptography, and secure hardware, represented by three mainstream technologies: federated learning, secure multi-party computation, and trusted execution environment. The underlying cryptographic technology consists of homomorphic encryption, differential privacy, secret sharing, and zero knowledge proof, and is closely integrated with blockchain technology to construct a complete technical system architecture (Wang et al., 2020). This approach minimises privacy risks by avoiding the transfer of raw data to a central server and helps overcome data silos. A typical federated learning system consists of multiple participants and a central server, each holding a local dataset with complete data features and almost no intersection between data samples. Federated learning was preferred over secure multi-party computation or trusted execution environments because it avoids the latency and hardware dependency of the latter while still keeping raw learner data on local devices, thus meeting institutional compliance policies without sacrificing predictive accuracy. Through the coordination of the central server, participants can join forces to train a more efficient global model. During the process of uploading private data, there is a high possibility of privacy data leakage. Both the transmission process and the data stored on the central server are not safe practices. Data transmitted beyond local systems constantly risks leakage, which traditional machine learning cannot secure (Wei et al., 2014).

In order to more accurately illustrate the implementation details of the federated learning framework, the following is the detailed process of the framework: assuming that K clients $\{Q_i\}_{i=1}^{K}$ want to complete data analysis, they all have their own private data $\{P_i\}_{i=1}^{K}$. The central server distributes the global model MFL to all clients $\{Q_i\}_{i=1}^{K}$ that need to be trained. After receiving the global model, the clients use their private data locally to train the global model, and can obtain their own local training models $\{M_i\}_{i=1}^{K}$. Unlike traditional machine learning, the clients only need to upload the obtained $\{M_i\}_{i=1}^{K}$ to the central server. When the central server receives the $\{M_i\}_{i=1}^{K}$ uploaded by all clients, thereby obtaining a new global model MFL and continuously loop this process, which will make the updates of *MFL* more and more accurate. Assuming the accuracy *VSUM* of the traditional machine learning model *MSUM* and the accuracy *VFL* of the federated learning model MFL, we can obtain:

$$|V_{FL} - V_{SUM}| < \delta \tag{12}$$

δ is a very small number, and the meaning of this formula is that it can approximate the accuracy of traditional machine learning and federated learning. If δ is small enough, it can indicate that the functions of traditional machine learning and federated learning are similar, and federated learning's lack of local training methods is sufficient to demonstrate its superiority. When the central server sends the model parameters w^t to all clients that require federated learning training data, upon receiving w^t, the clients perform local data training and obtain the trained local model:

$$F_i\left(w^t\right) = \frac{1}{|P_i|}\sum\nolimits_{u \in P_i} f\left(w^t, u\right) \tag{13}$$

in this process, f represents the loss function, and P_i represents the privacy data of the i^{th} client. The client can obtain the gradient of the local model based on the obtained local model $F_i(w^t)$:

$$g_i\left(w^t\right) = \frac{1}{|P_i|}\sum\nolimits_{u \in P_i} \nabla f\left(w^t, u\right) \tag{14}$$

all clients upload the gradients $g_i(w^t)$ of the local models to the central server. The central server summarises all the gradients of the local models received and then optimises the federated learning model *MFL*.

$$w^* = \arg\min F(w) \tag{15}$$

$$F(w) = \frac{1}{K}\sum\nolimits_{i=1}^{K} F_i(w) \tag{16}$$

$$w^{t+1} = w^t - \frac{\eta}{K}\sum\nolimits_{i \in K} g_i\left(w^t\right) \tag{17}$$

w^{t+1} represents the optimised federated learning model, w^* is the corresponding parameter, and η represents the learning rate. The K clients $\{Q_i\}_{i=1}^{K}$ mentioned above each have their own privacy data $\{P_i\}_{i=1}^{K}$, where P_i is a privacy dataset consisting of three parts (I, X, Y): one part is data denoted as I, another part is feature denoted as X, and the last part is label denoted as Y. Data I, feature X and label Y form a client dataset denoted as $P_i = (I, X, Y)$.

3 Framework of computer course path optimisation model

3.1 Learning path generation

Personalised learning path is a structure for organising learning objects, which not only involves the content to be learned, but also the order in which they are learned. The continuity relationship between research objects is an important feature factor that needs to be considered when optimising learning paths, and a reasonable sequence relationship is an important prerequisite for generating high-quality learning paths (Dornheim et al., 2022). For knowledge points, the continuity problem between them is the order in which they are learned. For example, in a course, one knowledge point may be the foundation or deepening of another knowledge point, and this order is constrained by the relationships between knowledge points, which are generally captured through knowledge models. The course knowledge graph is a graph structure consisting of numerous knowledge entities and relationships, denoted as $G = \{N, R\}$. Here, $N = (h, t)$ indicates the collection of knowledge entities within the graph, and $R = \{r|h, t \in N\}$ represents the relationships between nodes.

This article leverages the existing knowledge graph structure of computer courses to extract a subgraph of knowledge points from the initial learning path resources. The relationships between these points indicate their sequential order. The extracted knowledge point graph, derived from knowledge graph G_K, consists of nodes representing knowledge points within the initial learning path resources, with edges denoting sequential relationships. The adopted subgraph was mined from the publicly curated MOOCCube schema, whose directed prerequisite links among Java topics form a lightweight, acyclic backbone that enforces pedagogical coherence and keeps federated optimisation within feasible complexity bounds. The initial learning path of a computer course is denoted as $RLP = \{r_1, r_2, \ldots, r_{|RLP|}\}$, where each node r_i contains multiple knowledge points represented as $r_i = \{k_1, k_2, \ldots, k_{|ri|}\}$. The knowledge point subgraph extraction process is outlined as follows:

1 for each node r_i in the initial learning path RLP, search for knowledge point entities in the course knowledge graph G. If $k_i \in r_i$, store k_i in the knowledge point graph G_K

2 for the knowledge point entities k_i and k_j in G_K, if there exists a triplet relationship in the knowledge graph G such as (k_i, *relation*, k_j), then add the relation to G_K

3 repeat steps 1 and 2 until all nodes in the initial learning path RLP have been traversed

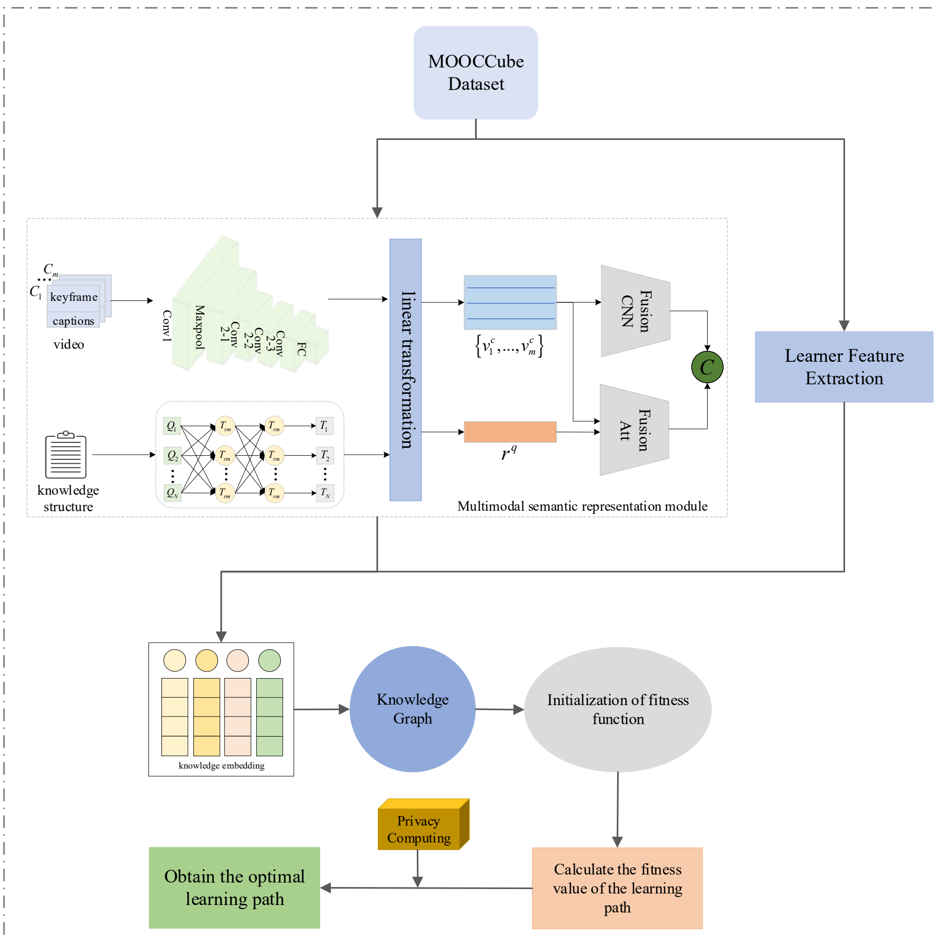
4 output the extracted knowledge point graph $G_K = \{K, Rel\}$, where $K = \{k_1, k_2, \ldots, k_{|K|}\}$, and $K \in N$ represents the set of extracted knowledge point entities, $Rel = \{relation_1, relation_2, \ldots, relation_{|Rel|}\}$ and $Rel \in R$ represents the set of relationships between the extracted knowledge points.

Based on the above extraction process, this paper studies the optimisation problem of learning path in the scenario. The optimisation process is as follows: given the subgraph structure G_K composed of several knowledge points that meet the learner's preferences in the initial learning path, the learner first selects knowledge points without preceding nodes from the knowledge points contained in the first predicted resource as the initial learning points, and then traverses and learns according to the existing paths in the knowledge point graph. After learning all the nodes in the knowledge point graph, the learner ends the learning process, where the nodes in the subgraph can only learn once. Therefore, the ultimate result of the above problem is to output a learning path composed of knowledge points, which is a comprehensive matching of knowledge features and learner personalised features.

3.2 MPCO model

To address the limitations of traditional computer course learning path planning, this paper proposes a multi-modal privacy collaborative learning path optimisation model MPCO. This model integrates multi-modal large models with privacy computing technology, aiming to optimise learning path planning. The overall architecture of the model is divided into three core parts: multi-modal data acquisition and preprocessing module, multi-modal large model learning and analysis module, and privacy computing protection module, as shown in Figure 1.

Figure 1 MPCO model framework diagram (see online version for colours)



In the multi-modal data acquisition and preprocessing module, the model integrates multiple data acquisition interfaces to collect various types of data from learners during the computer course learning process, including but not limited to text assignments, code practice results, graphic design works, online learning behaviour records, and stage test scores, among other heterogeneous data sources. These raw data undergo preprocessing processes such as denoising, standardisation, and standardised format conversion to ensure data quality and compatibility, enabling them to be used as effective inputs for subsequent model stages.

The multi-modal large model learning and analysis module is the core of the model. Based on preprocessed data, a multi-modal large model using transformer architecture is used for deep learning and analysis of learners' multidimensional features. The model captures the correlation and fusion features between different modal data such as text, images, and code through self-attention mechanism, and constructs a comprehensive and accurate learner portrait. This portrait covers key information such as the distribution of learners' knowledge mastery, learning preference styles, cognitive ability levels, and potential learning difficulties, providing data support for personalised learning path planning.

The privacy computing protection module runs through the entire model operation process. In the data storage stage, homomorphic encryption technology is used to encrypt the original learning data, ensuring the security of the data in a static state. When performing computational tasks such as model training or learning path planning, differential privacy techniques add an appropriate amount of noise to the data to prevent the analysis results from inferring sensitive personal information. By tightly coupling these privacy measures with the multi-modal large model, MPCO gains access to far richer, cross-institutional datasets than any single-modal or unsecured approach, yielding both higher predictive accuracy and strict confidentiality.

The above model framework design combines the powerful learning and analysis capabilities of multi-modal large models with strict data security guarantees for privacy computing, aiming to provide accurate, personalised, and secure learning path optimisation solutions for computer course learning. The optimisation model for computer course learning paths in this article must satisfy certain conditions, which are translated into four objective functions F_1 to F_4. Their mapping relationship F is depicted in equation (18):

$$F = \begin{cases} F_1(X) = \sqrt{\sum_{j=1}^{|K|} \left| \dfrac{\sum_{i=1}^{|K|} \left[X_{ij}\left(D_i - A_i\right) + X_{ij}\left(D_j - A_j\right) \right]}{2\sum_{i=1}^{|K|} X_{ij}} \right|^2} \\ F_2(X) = \dfrac{\sum_{j=1}^{|K|} \sum_{i=1}^{|K|} X_{ij} \left| I_i - I_j \right|}{\sum_{j=1}^{|K|} \sum_{i=1}^{|K|} X_{ij}} \\ F_3(X) = \sum_{j=1}^{|K|} \sum_{i=1}^{|K|} X_{ij} S_{ij} \\ F_4(X) = \sum_{j=1}^{|K|} \sum_{i=1}^{|K|} X_{ij} H_{ij} \end{cases} \tag{18}$$

The four objective functions $F_i(X)$ are as follows: $F_1(X)$ measures the difficulty gap between learning path knowledge points and the learner's mastery level; $F_2(X)$ ensures a balanced importance of knowledge points along the path; $F_3(X)$ calculates the total learning cost for the recommended path's knowledge points; and $F_4(X)$ evaluates the overall learning experience of the recommended path.

4 Experimental results and analyses

This article uses computer course data from the publicly available dataset MOOCCube as experimental data. The learner behaviour data and course data in the dataset are from Xuetang.com. The student behaviour data mainly includes the number of video views, total video duration, actual viewing time, video playback time, starting position, ending position, earliest starting time, and latest ending time of the user. In order to use learners' learning behaviour data for the experiment in this article, it is necessary to first preprocess the obtained data and convert it into a dataset that fits the research scenario in this article. Firstly, the behaviour data of each learner is sorted according to their starting time of learning, and then the 'u_id' and 'video_id' in the dataset are renumbered, where 'video_id' is arranged according to the order in which the video appears in the original course outline. Normalise the learner interaction data, filter out users with only one interaction behaviour and redundant data, and finally convert the remaining data into a dataset that meets the research scenario of this article according to experimental requirements. The selected pipeline – temporal reordering, user/video re-indexing, z-score normalisation and singleton filtering – was designed to suppress timestamp drift, align sparse interaction logs to the transformer's fixed-length input, and balance class frequencies, thereby sharpening signal-to-noise ratios without inflating computational cost.

In order to verify the effectiveness of the proposed multi-modal large model and privacy computing collaborative driven computer course learning path optimisation model MPCO, this paper considers the following baseline methods for comparison:

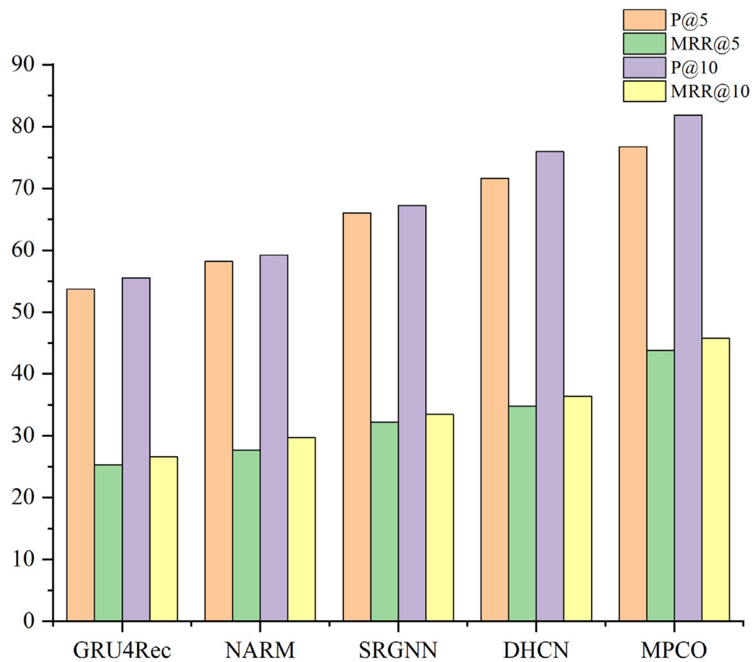
- GRU4Rec: using RNN neural network to process long-term features of sequences
- NARM: combining attention mechanism with RNN to highlight the importance of items
- SRGNN: first, the sequence is constructed into a graph, and object representation is learned by propagating information on the graph
- DHCN: sequence recommendation based on self-supervised hypergraph convolutional network.

For the sake of fairness, this article sets the vector length of all models to 100, the batch size of all models to 100, $L2$ regularisation to 10-5, and an initial learning rate of 0.001. For GNN-based methods, this article searched for {1, 2, 3, 4} layers.

Table 1 Performance comparison of baselines

Model	*P@5*	*MRR@5*	*P@10*	*MRR@10*
GRU4Rec	53.71	25.32	55.53	26.56
NARM	58.15	27.63	59.24	29.68
SRGNN	66.03	32.16	67.22	33.48
DHCN	71.62	34.81	75.93	36.38
MPCO	*76.75**	*43.79**	*81.82**	*45.74**

Figure 2 Performance comparison of baselines (see online version for colours)



The baseline methods selected in this article mainly include three categories: RNN-based methods (GRU4Rec and NARM), GNN-based methods (SRGNN), and hypergraph neural network-based methods (DHCN). By comparing the experimental results of the MPCO model in this article with the latest methods mentioned above, we aim to verify the comprehensive performance of our model in optimising the learning path of computer courses. From the experimental results in Table 1, it can be seen that in the RNN-based method, the NARM model captures users' long-term interest preferences by utilising attention mechanisms, and its experimental results are significantly better than the early GRU4Rec method. This indicates that introducing users' long-term preferences is very important in the task of learning path prediction. Compared with RNN-based methods, GNN-based methods exhibit better performance. In the SRGNN model, not only is the user's behaviour interaction sequence used, but also the topological information of the graph structure is utilised, so the model can more comprehensively represent the characteristics of projects and events. In addition, as a suboptimal result, the DHCN model constructs an undirected sequence hypergraph structure when modelling user behaviour sequences, and uses a hypergraph convolutional network (HGCN) to represent the features of the hypergraph. From the results, it can be seen that the DHCN model is significantly superior to other methods, indicating the need to construct a hypergraph structure to represent the high-order information hidden between behavioural data. Compared with the latest baseline methods mentioned above, the MPCO model proposed in this paper exhibits better predictive performance. This superior performance stems from MPCO's unique combination of encrypted federated learning, multi-modal fusion, and temporal self-attention – three elements that together reveal nuanced learner trajectories and evolving preferences that RNN, GNN or hypergraph baselines cannot jointly capture. Unlike prior single-modal or shallow-fusion systems, MPCO employs a transformer-based multi-head attention mechanism that jointly embeds textual, visual and code artefacts into a unified latent space, capturing subtle cross-modal interactions that yield markedly richer and more discriminative learner portraits. The modelling method in this article integrates multiple heterogeneous data sources such as text assignments, code practices, and graphic designs through a multi-modal large model, comprehensively mining the multidimensional features of learners and constructing accurate learner portraits. At the same time, utilising privacy computing technology to ensure the security of learning data, eliminating data silos, and achieving secure sharing and effective utilisation of data. On this basis, the MPCO model captures learners' dynamic learning preferences by introducing a self-attention mechanism with temporal position signals, which can more accurately predict the most suitable resources for learners to learn in the next step, thereby optimising the learning path. Unlike vanilla attention, the temporal position-aware variant explicitly encodes both inter-event intervals and sequential order, allowing the model to detect forgetting curves and recency biases that static attention weights overlook. The experimental results show that the predictive performance of the MPCO model in learning path optimisation significantly exceeds all baseline methods, verifying the effectiveness of the proposed multi-modal large model and privacy computation driven learning path optimisation method in this paper.

To evaluate how multi-modal large models and privacy computing affect our learning path optimisation model, we have created three model MPCO variants. The first two variants focus on assessing the impact of multi-modal large models, while the third one aims to test the effectiveness of privacy computing within the model:

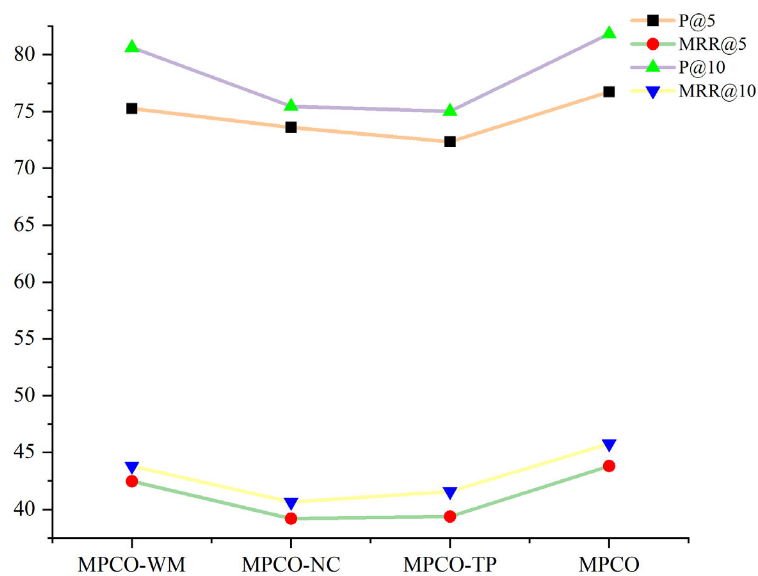
- MPCO-WM: in the MPCO method, we only use single modal data to embed node features, comparing the experimental results of multimodal and single modal approaches
- MPCO-NC: in the MPCO method, we do not use privacy computation to encrypt learner data, that is, to verify the improvement of model performance by introducing privacy computation
- MPCO-TP: in the MPCO method, we did not use a self-attention mechanism with temporal position signals to validate the effectiveness of dynamic learning preferences in the model.

Table 2 Performance comparison of ablation study

Model	*P@5*	*MRR@5*	*P@10*	*MRR@10*
MPCO-WM	75.28	42.46	80.63	43.79
MPCO-NC	73.62	39.21	75.47	40.64
MPCO-TP	72.36	39.38	75.04	41.58
MPCO	*76.75**	*43.79**	*81.82**	*45.74**

It can be seen from the ablation experiment results in Table 2 that the overall learning path prediction performance of the model MPCO-WM using only single modal data is lower than that of the original MPCO model. The multi-modal data fusion method outlined in this paper substantially enhances model performance optimisation. Multi-modal data offers more comprehensive learner feature information, reduces many error prone paths from single modal limitations, and subsequently improves model prediction. For the model MPCO-NC that does not use privacy computing for data encryption, its prediction effect is not only lower than the original model MPCO, but also lower than the model MPCO-WM that only uses single modal data. This is because in the MPCO-NC model, the lack of privacy computation protection may lead to data security issues, resulting in more erroneous results during prediction. At the same time, the absence of privacy computation may also affect the integrity and availability of learner data. From the experimental results of the two variant models mentioned above, it can be seen that the multi-modal data fusion and privacy computing techniques introduced in this paper for modelling learning behaviour help improve the predictive performance of the model, and the overall performance improvement effect of the model is significant. Compared to the original model MPCO, the model MPCO-TP, which lacks the self-attention mechanism with temporal position signals, shows significantly lower predictive performance. In MPCO-TP, removing this mechanism forces the model to rely solely on static learner preferences for predictions. Experimental results indicate that incorporating the self-attention mechanism with temporal and positional signals enhances the model's predictive capability. This underscores the practicality of using dynamic learning preferences, aligning better with actual learner needs and preferences.

Figure 3 Performance comparison of baselines (see online version for colours)



In the experimental design of this study, a case study involving Java programming language learning path optimisation was conducted. The case focused on a group of 50 novice Java learners from a university's introductory computer science course. The MPCO model was applied to optimise their learning paths based on multi-modal data collected during an eight-week semester. The data included weekly Java coding assignments, results from online Java concept quizzes, and records of their interactions with Java learning materials such as video lectures and documentation. The model analysed this multi-source information to identify each learner's strengths and weaknesses in Java fundamentals like syntax, object-oriented programming concepts, and basic algorithms. For instance, the model detected that some learners struggled with understanding polymorphism and inheritance despite performing well in syntax-related tasks. Based on these insights, the MPCO model generated personalised learning paths that recommended specific Java learning resources. For learners showing weaknesses in object-oriented principles, the system suggested targeted tutorials and practice exercises on classes, objects, and inheritance. The optimised paths also incorporated privacy computing by ensuring that all learner data, including their quiz scores and coding assignment details, were encrypted during storage and processing. This case demonstrated how the integration of multi-modal analysis and privacy protection could enhance the learning experience. Learners following the MPCO-optimised paths showed a 28% improvement in their final Java project scores compared to those on traditional learning paths. This practical example underscores the model's potential to address real-world challenges in computer course learning path optimisation.

5 Conclusions

In this article, we delve into the application of multi-modal large models in personalised learning path planning and propose an innovative learning path optimisation scheme by combining privacy computing technology. In the experimental stage, we adopted a comparative experiment method to compare and analyse different algorithms to verify the effectiveness of the proposed method. The experimental results demonstrate that the method based on multi-modal large models and privacy computing outperforms traditional single-modal or non-privacy-protecting methods in accuracy, adaptability, and security for learning path planning. By optimising computer course learning paths, we have found this approach effectively boosts learners' efficiency. Empirical research indicates that enhanced learning efficiency allows learners to grasp course content and finish tasks more quickly, with significant improvement in knowledge mastery. This shows that learning path planning optimisation focuses not only on learning speed but also on quality and effectiveness. Moreover, the research highlights the significance of personalised learning experiences. By adapting to learners' progress and interests, this optimisation strategy offers a more personalised learning experience, sparking learners' interest and enthusiasm. Such personalised learning path planning meets diverse learner needs, increasing their satisfaction and sense of achievement. In summary, this study offers new insights and methods for personalised education development, particularly in the collaborative use of multi-modal large models and privacy computing. Future work could apply this strategy to more educational scenarios, providing learners with more efficient, secure, and personalised learning path planning.

Declarations

The author declares that he has no conflicts of interest.

References

Dornheim, J., Morand, L., Zeitvogel, S., Iraki, T., Link, N. and Helm, D. (2022) 'Deep reinforcement learning methods for structure-guided processing path optimization', *Journal of Intelligent Manufacturing*, Vol. 33, pp.333–352.

Dwivedi, P., Kant, V. and Bharadwaj, K.K. (2018) 'Learning path recommendation based on modified variable length genetic algorithm', *Education and Information Technologies*, Vol. 23, pp.819–836.

Goyal, B., Dogra, A., Lepcha, D.C., Koundal, D., Alhudhaif, A., Alenezi, F. and Althubiti, S.A. (2022) 'Multi-modality image fusion for medical assistive technology management based on hybrid domain filtering', *Expert Systems with Applications*, Vol. 209, p.118283.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C. and Xu, Y. (2022) 'A survey on vision transformer', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 1, pp.87–110.

Huang, X.J., Zhang, C. and Zheng, Q.H. (2014) 'Generating personalized navigation learning path based on knowledge map', *International Journal of Technology and Educational Marketing*, Vol. 4, No. 2, pp.1–17.

Jiang, B., Li, X., Yang, S., Kong, Y., Cheng, W., Hao, C. and Lin, Q. (2022) 'Data-driven personalized learning path planning based on cognitive diagnostic assessments in MOOCs', *Applied Sciences*, Vol. 12, No. 8, p.3982.

Li, F., Li, H., Niu, B. and Chen, J. (2019) 'Privacy computing: concept, computing framework, and future development trends', *Engineering*, Vol. 5, No. 6, pp.1179–1192.

Liu, H. and Li, X. (2020) 'Learning path combination recommendation based on the learning networks', *Soft Computing*, Vol. 24, No. 6, pp.4427–4439.

Liu, S., Gao, P., Li, Y., Fu, W. and Ding, W. (2023) 'Multi-modal fusion network with complementarity and importance for emotion recognition', *Information Sciences*, Vol. 619, pp.679–694.

Nabizadeh, A.H., Goncalves, D., Gama, S., Jorge, J. and Rafsanjani, H.N. (2020) 'Adaptive learning path recommender approach using auxiliary learning objects', *Computers & Education*, Vol. 147, p.103777.

Niknam, M. and Thulasiraman, P. (2020) 'LPR: a bio-inspired intelligent learning path recommendation system based on meaningful learning theory', *Education and Information Technologies*, Vol. 25, No. 5, pp.3797–3819.

Shmelev, V., Karpova, M. and Dukhanov, A. (2015) 'An approach of learning path sequencing based on revised Bloom's taxonomy and domain ontologies with the use of genetic algorithms', *Procedia Computer Science*, Vol. 66, pp.711–719.

Tan, W., Thitøn, W., Xiang, P. and Zhou, H. (2021) 'Multi-modal brain image fusion based on multi-level edge-preserving filtering', *Biomedical Signal Processing and Control*, Vol. 64, p.102280.

Tetko, I.V., Karpov, P., van Deursen, R. and Godin, G. (2020) 'State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis', *Nature communications*, Vol. 11, No. 1, p.5575.

Vanitha, V., Krishnan, P. and Elakkiya, R. (2019) 'Collaborative optimization algorithm for learning path construction in e-learning', *Computers & Electrical Engineering*, Vol. 77, pp.325–338.

Wang, J. (2025) 'Optimizing learning path design in mobile learning platforms for online courses', *International Journal of Interact. Mob. Technol.*, Vol. 19, No. 1, pp.4–17.

Wang, T., Mei, Y., Jia, W., Zheng, X., Wang, G. and Xie, M. (2020) 'Edge-based differential privacy computing for sensor-cloud systems', *Journal of Parallel and Distributed Computing*, Vol. 136, pp.75–85.

Wei, L., Zhu, H., Cao, Z., Dong, X., Jia, W., Chen, Y. and Vasilakos, A.V. (2014) 'Security and privacy for storage and computation in cloud computing', *Information Sciences*, Vol. 258, pp.371–386.

Zhang, X., Gong, Y., Lu, J., Wu, J., Li, Z., Jin, D. and Li, J. (2023) 'Multi-modal fusion technology based on vehicle information: a survey', *IEEE Transactions on Intelligent Vehicles*, Vol. 8, No. 6, pp.3605–3619.

Zhao, J., Mao, H., Mao, P. and Hao, J. (2024) 'Learning path planning methods based on learning path variability and ant colony optimization', *Systems and Soft Computing*, Vol. 6, p.200091.

Zhou, Y., Huang, C., Hu, Q., Zhu, J. and Tang, Y. (2018) 'Personalized learning full-path recommendation model based on LSTM neural networks', *Information Sciences*, Vol. 444, pp.135–152.

Zhu, H., Tian, F., Wu, K., Shah, N., Chen, Y., Ni, Y., Zhang, X., Chao, K-M. and Zheng, Q. (2018) 'A multi-constraint learning path recommendation algorithm based on knowledge map', *Knowledge-Based Systems*, Vol. 143, pp.102–114.