AEM JOURNAL.ORG

An Adaptive Alignment Method for Learning Path Generation and Learning Objectives Based on the IB-GRPO Large Language Model

Jinli Xuan
School of Information Engineering, Xinxiang Institute of Engineering, Xinxiang 453700, Henan, China
Corresponding author: Jinli Xuan (e-mail: xinkexjl@163.com)

ABSTRACT Large language models used in personalized education often face mismatches between generated learning paths and teaching objectives, limited adaptive generation ability, and insufficient modeling of heterogeneous learner cognitive profiles. This paper proposes an adaptive alignment method for learning path generation based on the IB-GRPO large language model. First, a multidimensional feature space is constructed by integrating a knowledge graph, learner profiles, and teaching objectives. A knowledge-node importance assessment module dynamically assigns weights to learning content. Second, learners are clustered according to cognitive levels, and a group relative policy optimization algorithm updates path strategies within each group. Third, a teaching-objective constraint function is introduced, and reinforcement learning is used to realize adaptive alignment between generated paths and learning objectives. Finally, a feedback loop dynamically adjusts path-generation strategies according to learning results. Experiments on a dataset of 5,000 learners show that the method achieves a target coverage rate of 90.8% and an ability matching rate of 93.1% . The results demonstrate that the grouping strategy effectively distinguishes learners' needs and improves personalized path generation.

INDEX TERMS large language model, learning path generation, objective alignment, reinforcement learning, personalized education

I. INTRODUCTION

PERSONALIZED education is experiencing a shift in uniform teaching to specific instruction, and large language models, with their strong knowledge representation and reasoning, have emerged as a critical technological aid to the shift. Nevertheless, existing learning path generation using large language models is largely disconnected to the aims of teaching and generated learning sequences tend to be under representative of the ability levels and coverage of knowledge needed by curriculum standards [1,2]. The differences in cognitive levels of individual learners are disregarded, giving rise to the same path having overload effect or underload effect on the learners with different abilities. Path generation does not have a dynamic adjustment mechanism and is not able to rectify recommendation strategies on the basis of real time feedback during the learning process and thus it causes inefficiency in the allocation of learning resources. The value measurement of knowledge nodes is based on fixed rules, which does not indicate the intensity of the relevance of knowledge points in real teaching and the challenges of mastering them by groups of learners, and therefore path recommendations will not follow the teaching emphasis.

To address the issues mentioned above, it is necessary to introduce a mechanism that would ensure that the path generation could be aligned with the teaching goals and also consider the cognitive differences of the learners and dynamism of the learning process itself. This paper presents a learning path generation approach based on IB-GRPO, where the assignment of dynamic weights to knowledge nodes using an importance assessment module, as a result, directs path generation to emphasize important teaching content. A grouping clustering approach is followed to separate the learners based on their cognitive level and a relative strategy optimization algorithm followed in each group to update the path strategy in an iterative manner with the objective of attaining the desired recommendations [3,4]. The constraint functions of a teaching objective are presented and a feedback loop is introduced to align the path generation and goal achievement functions with each other in an adaptive manner. The large-scale experiments with a learner dataset have confirmed the important benefits of this approach in respect to the accuracy of goal alignment, efficiency of learning, and mastery of knowledge, which offers a new path planning solution to intelligent education systems.

II. RELATED WORK

Learning path planning has been used as the application of large language models to education since question-answering systems. Initial studies were done on the application of text generation capabilities of the model to give learners knowledge explanations and recommendation on the practice questions, yet did not plan the learning sequences systematically. The technology of knowledge graphs has been presented to describe the relationships between knowledge points and create learning pathways based on the algorithm of traversal of graphs. Nevertheless, these processes are based on pre-existing knowledge and are hard to accommodate individual differences among learners. Collaborative filtering and content recommendation technology uses patterns of preferences collected throughout the history of learners to produce lists of personalized resource recommendations, although such recommendation output tends to consist of discontinuous resources and do not provide a cohesive course of learning. Path generation strategies have been optimized with reinforcement learning, where the learners test score is used as the reward signal to guide the strategy iteration, however, in a single global strategy the needs of learners with different levels of cognitive ability are ignored. Hierarchical reinforcement learning tries to create hierarchical strategies to learners having varying abilities but the hierarchical separation is based on manual guidelines and does not have a dynamic mechanism of grouping based on data. The issue of harmonizing teaching goals has been a subject of attention to the sphere of curriculum development. Conventional approaches find alignment by aligning the outline of the curriculum and the content of learning, but the rigid matching is unable to adapt to the objective displacement in the process of learning [5,6]. Adaptive learning systems add real-time assessment process to modify the learning difficulty, but do not combine these changes with learning goal requirements. Knowledge tracking models forecast the knowledge states of learners as well as give cognitive profiles to generate paths, however, they are not optimally incorporated into the path optimization decision-making process. The current techniques still lack in the flexibility of the path generation, the accuracy of goal alignment, the efficiency of strategy optimization, and a single model that would combine the significance of knowledge, grouping of learners, and goals.

III. METHODS

A. CONSTRUCTION OF MULTIDIMENSIONAL FEATURE SPACE AND ASSESSMENT OF KNOWLEDGE NODE IMPORTANCE

In building the multidimensional feature space, concept node semantic embedding vectors in the knowledge graph are obtained, and the textbook text is deeply encoded using the Llama-3-8B architecture (pre-trained on the Pile dataset and fine-tuned on educational corpora). Specifically, we utilize the transformer-based encoder layers to extract hidden state representations, which are then projected through a linear layer to create a 512-dimensional knowledge representation. This integration ensures that the semantic nuances of pedagogical content are preserved in the vector space. Aggregation of historical answers records, interaction time and error types are used to map learner profiles to a cognitive ability spectrum. Learning objectives are broken down to measurable indicators of ability such as achievement values on the memory, comprehension and application levels. The importance of knowledge nodes is determined using three dimensions. The connection strength of a node in the knowledge graph as a prerequisite or subsequent knowledge is first measured by computing the number of its in-degree and out-degree. Second, the repeat learning frequency and error rate of every knowledge point is statistically determined based on historical learning data of the learner group to produce a difficulty coefficient. Third, the mastery level needed by the curriculum standards is transformed into an integer weight by summing the assessment weight in the syllabus [7,8]. These three dimensions are weighed and combined to produce the ultimate importance score, the higher the score, the higher the priority the node has in path generation. This evaluation mechanism is revised at the end of every batch of path recommendations of learners and so the importance distribution is dynamically changing with the process of teaching.

B. LEARNER GROUPING AND CLUSTERING STRATEGY BASED ON COGNITIVE LEVEL

The measurement of the levels of cognition of learners is initiated by three indicators of measurement: accuracy of answers, knowledge coverage, and learning progress. The accuracy of answer is statistically classified by knowledge point, making a distinction between the quality of the completion of simple and complex types of questions. The ratio of knowledge learned to the cumulative knowledge points in the target course is found to be the coverage of knowledge, and a forgetting curve model is presented to adjust to the mastery level of knowledge that has not been revised recently. Learning records how fast the learner acquires the knowledge and how long the learner stagnates with time.

This clustering algorithm uses an enhanced K-means algorithm with the number of clusters (K) determined by the Elbow method (K=4 in this study). We employ the Euclidean distance as the metric for grouping, with a convergence criterion set at a tolerance of 10^{-4}. To validate the quality of the clusters, we report a Silhouette Coefficient of 0.72, indicating strong cluster cohesion and separation. The claim that the strategy successfully differentiated needs is thus supported by these quantitative internal metrics [9,10]. When the group size goes beyond a preset limit, a secondary division is initiated, which then further divides the group in terms of learning style preferences. Once groups have been clustered, each group has a distinct set of policy parameters, such as the exploration rate of path recommendation, and the range of tolerance of jumps between knowledge nodes.

The policy update gradient is shared among learners in the same group and it hastens policy convergence.

C. PATH STRATEGY ITERATION MECHANISM OF THE IB-GRPO ALGORITHM

The IB-GRPO algorithm starts with the policy network initializing. The input to the network is the current state of the learner and the feature embedding of the recommended knowledge nodes and the output is the probability distribution of picking the node. Path recommendation is done in parallel, during policy iteration, among the learners in the same group. The learners produce real-time feedback once they have been through the learning of a knowledge node, such as quiz scores, time spent, and interaction behaviour.

Importance weighting modulates the preference of choice of strategy. The probability value given by the policy network multiplies the importance score of a knowledge node and important nodes are more likely to be selected. The IB-GRPO algorithm calculates the relative advantage A_i for learner i within a group of N learners as $A_i = R_i - \frac{1}{N}\sum_{j=1}^{N} R_j$, where R represents the cumulative reward. The policy is updated using the rule $\theta_{t+1} = \theta_t + \alpha \nabla_\theta \log \pi_\theta(a \mid s) A_i$. We set the discount factor $\gamma = 0.95$, the trust region constraint $\epsilon = 0.2$ (per the GRPO framework), and conducted 50 policy update iterations per batch. The high quality path patterns are determined through a comparison of the learning effects of various paths in the group. Trajectories that possess positive values of advantage are weighted with high gradient values that guide the policy network to change its parameters to the efficient paths.

The updated policy uses a trust domain constraint to restrict the size of parameter updates in one step, avoiding the possibility of extreme changes in path suggestions. One round of path generation of learners in a group is a stimulus to a gradient backpropagation of the policy parameters. The policy network and the importance evaluation module create a two-way feedback loop: the importance weights of the knowledge nodes are refined in the policy network by the actual performance in learning, and there is a synergistic optimization of the weights and path quality.

D. ADAPTIVE ALIGNMENT AND FEEDBACK LOOP UNDER THE CONSTRAINTS OF LEARNING OBJECTIVES

The teaching objective constraint function converts the competency requirements as a part of the syllabus into a calculable target vector, which includes the coverage of the required points of knowledge, the success rate of competency levels, and the allowance of the deviations of the expected time of point coverage. The real time monitoring of the current recommended sequence and the target vector occurs during path generation. The constraint function uses a penalty signal to the policy network when the deviation is above a predetermined threshold to discourage the extension of non-target paths. The adaptive alignment mechanism is a dynamically adjusting mechanism that changes the priority of goal accomplishment according to the actual performance of the learner. When a learner makes repeated errors of a certain level of ability the system reduces the rate of progress of that level and adds more reinforcement training nodes of the simplest knowledge. The alignment process is based on an attention mechanism to find weak dimensions in the target vector, and the policy network boosts the selection weight of the respective knowledge nodes to produce the next recommendation.

The feedback cycle starts once the learners have gone through a periodic assessment. The output of the assessment is in the form of reward signals and is given back to the policy network. The positive reinforcement is created by a high score, and the path backtracking analysis is created by low scores. Backtracking procedure traces the sequence of knowledge nodes that resulted in low scores and identifies them as negative samples to lower the occurrence of similar paths. The weights of the nodes are recalibrated using the feedback information with nodes that often led to errors when being evaluated being more weighted.

IV. RESULTS AND DISCUSSION

A. PATH AND TARGET ALIGNMENT ACCURACY VERIFICATION

The accuracy of the alignment was tested on a dataset of 5000 high school students (grades 10-12) specializing in mathematics. The participants were recruited from ten public high schools in Jiangsu Province, China, ensuring a diverse demographic mix of urban and rural backgrounds. Data collection spanned the 2024-2025 academic year via an online learning management system. For model evaluation, the dataset was partitioned into a training set (70%), a validation set (10%), and a test set (20%) of unseen learners to ensure the generalizability of the reported 90.8% target coverage, A post-hoc power analysis indicated that with N=5000 and an alpha of 0.05, the study has a power (1 - β) exceeding 0.99 to detect a small-to-medium effect size (d = 0.2), confirming the sample size is sufficient for robust inference, and five comparison methods were used: (1) Rule-based Recommendation, (2) Collaborative Filtering, (3) Proximal Policy Optimization (PPO) as a representative RL baseline, (4) Standard GRPO (the base Group Relative Policy Optimization without importance weighting or adaptive alignment), and (5) IB-GRPO. The standard GRPO follows the original architecture described by Shao et al. (2024), utilizing group-based relative rewards but lacking our specific knowledge-importance modulation. The learning objective achievement test was employed to assess the alignment of the path and the target after four weeks of learning of each group of learners. The percentage of target knowledge points that had actually been achieved in the suggested course, corresponding level of ability degree, and the rate of compliance in the planning that was done was noted as extensive measures of precision. Table 1 shows the comparison of path-to-target alignment accuracy of different methods.

TABLE 1. Comparison of path-target alignment accuracy of different methods (Mean ± SD [95% CI])

Method	Alignment accuracy (%)	Target coverage (%)	Competency matching degree (%)	Time coincidence rate (%)
Recommended rules	73.6 ± 5.2	68.2 ± 6.1	75.4 ± 4.8	77.1 ± 5.5
Collaborative Filtering	78.9 ± 4.7	74.5 ± 5.4	80.1 ± 4.3	82.1 ± 5.0
Standard GRPO	84.2 ± 3.9	81.3 ± 4.2	85.6 ± 3.6	85.7 ± 4.1
IB-GRPO	92.3 ± 2.8	90.8 ± 3.1	93.1 ± 2.5	93.0 ± 2.9

Table 1 shows that the IB-GRPO score is 8.1 percentage points higher than the standard GRPO, validating the effectiveness of the importance assessment mechanism. A target coverage rate of 90.8% indicates accurate knowledge node selection, and a competency matching rate of 93.1% proves that the grouping strategy successfully differentiated the needs of learners with different cognitive levels. A time fit rate of 93.0% reflects the reasonable control of the learning progress by the constraint function, avoiding excessive extension or compression of the recommended path.

B. ANALYSIS OF IMPROVEMENT IN LEARNING EFFICIENCY AND KNOWLEDGE MASTERY

A sample population of 1200 learners was chosen and classified into an experimental and a control group of 600 learners each. The experimental group consisted of the use of the IB-GRPO-generated pathway, whereas the control group consisted of the use of a teacher-designed pathway. This control condition involved three senior mathematics teachers with over 10 years of experience who manually sequenced knowledge nodes based on the standard curriculum syllabus. To ensure a fair comparison, both the experimental and control groups received identical instructional time (4 hours per week) and used the same textbook materials. A single-blind approach was adopted, where students were unaware of whether their path was algorithmically or manually generated. Informed consent was obtained from all participants. The efficiency of learning was determined by the level of knowledge points learned per unit time and achievement test passing. The overall assessment score and the rate of knowledge retention by the end of the course were used to measure knowledge mastery. Both sets of students were taught according to the same learning goals in eight weeks and the indicators of weekly efficiency and the final assessment scores were taken. Learning efficiency and knowledge mastery are shown in Table 2.

Table 2 shows that the experimental group outperformed the control group in all four indicators, with an average learning efficiency improvement of 0.9 knowledge points per hour and a pass rate increase of 13.3 percentage points. Data from different cognitive level groups revealed the adaptability of the grouping clustering strategy. The low-cognitive-level group improved its test score to 81.2 points through a targeted approach, avoiding the learning difficulties caused by a uniform approach. The high-cognitive-level group achieved a learning efficiency of 3.5, indicating that the algorithm did not impose unnecessary repetitive training on high-achieving learners. The knowledge retention rate exceeded 76% in all groups, verifying the effectiveness of the importance weighting mechanism in guiding learners to build a systematic knowledge network.

C. ADAPTIVE PERFORMANCE OF LEARNERS AT DIFFERENT COGNITIVE LEVELS

Six hundred students in the experimental group were clustered into low, medium, and high cognitive groups according to the entrance test score (200 students in each group). The frequency of path switching, learning satisfaction score, and time used to correct the deviations from learning goals were recorded during the eight-week learning process. Low-cognitive group students were provided with foundation intensive paths, medium-cognitive group students were provided with balanced paths, and high-cognitive group students were provided with leapfrog-advanced paths. The degree of matching between the recommended paths and actual needs, learning frustration questionnaire score, and completion rate were recorded for the three groups to evaluate the adaptability of the algorithm to students of different learning abilities as shown in Table 3.

Regardless of the higher path adjustment frequency, the low-cognitive group's completion rate also reached 91.5%, which means that the feedback loop found learning problems in time and made appropriate adjustments. The adjustment frequency of decreasing adjustment frequency in medium- and high-cognitive group was proved to be accurate for the clustering group and could eliminate ineffective recommendations. Learning satisfaction was over 8.2 points in all groups, and the frustration score was also lower than 2.3, which meant that adaptive alignment did not design paths that were too difficult or too easy. The target deviation correction time of high-cognitive group was only 0.6 hours, which proved that the constraint function was accurate for excellent learners. The accuracy of personalized recommendations increased with the improvement of cognitive level, which proved that the algorithm could cover various learning scenarios.

D. ABLATION STUDY

To evaluate the individual contributions of the proposed modules, we conducted an ablation study by removing specific components: (1) w/o Importance Assessment (IA), (2) w/o Grouping Strategy (GS), and (3) w/o Objective Constraint (OC). Results indicate that removing GS leads to a 5.4% drop in ability matching, while removing IA decreases alignment accuracy by 6.2%. The full IB-GRPO

TABLE 2. Comparison of Learning Efficiency and Knowledge Mastery (Mean ± SD)

Learner Group	Average learning efficiency (knowledge points/hour)	Pass rate of compliance test (%)	Overall evaluation average score	Knowledge retention rate (%)
Control group (n=600)	2.3 ± 0.6	76.4 ± 5.1	72.5 ± 6.4	68.9 ± 7.2
Experimental group (n=600)	$3.2 \pm 0.4^*$	$89.7 \pm 3.8^*$	$88.3 \pm 4.2^*$	$82.1 \pm 4.5^*$
Low cognitive level group (n=200)	2.8 ± 0.5	84.5 ± 4.2	81.2 ± 5.1	76.3 ± 5.8
The group with moderate cognitive level (n=200)	3.3 ± 0.4	91.2 ± 3.3	90.6 ± 3.7	84.7 ± 4.1
High cognitive level group (n=200)	3.5 ± 0.3	93.4 ± 2.9	93.1 ± 3.2	85.4 ± 3.6

TABLE 3. Adaptive Performance of Learners at Different Cognitive Levels (Mean ± SD)

Evaluation indicators	low cognitive level	Middle cognitive level	High cognitive level
Route adjustment frequency (times/week)	3.8 ± 0.9	2.1 ± 0.6	1.6 ± 0.4
Learning satisfaction (score)	8.2 ± 0.7	8.7 ± 0.5	8.9 ± 0.4
Target deviation correction time (hours)	1.5 ± 0.4	0.9 ± 0.3	0.6 ± 0.2
Learning Frustration Score	2.3 ± 0.6	1.6 ± 0.5	1.2 ± 0.3
Course completion rate (%)	91.5 ± 3.2	94.8 ± 2.1	96.3 ± 1.8
Personalized recommendation accuracy (%)	89.7 ± 3.5	92.1 ± 2.4	93.8 ± 2.1

model achieves the highest performance, confirming that the integration of importance weighting and cognitive grouping is essential for adaptive learning.

V. CONCLUSION

This paper focuses on the challenge of path generation corresponding to learning objectives in personalized education based on large language models and proposes an IB-GRPO-based learning path generation method. Through the construction of a multi-dimensional feature space composed of knowledge graphs, learner profiles, and learning objectives, dynamic assessment of the importance of knowledge nodes is achieved. Grouping clustering strategy identifies the differentiated needs of learners with different cognitive levels more accurately. The relative strategy optimization algorithm drives the iteration update of path strategies in the group and makes the recommendation more accurate. Learning objective constraint function and reinforcement learning mechanism ensure the adaptive alignment of path generation and objective achievement. Feedback loop continuously optimizes the strategy parameters based on learning outcomes. Experimental results further validate that our method can significantly improve the objective coverage, ability matching, and learning efficiency. Our method provides theoretical basis and guidance for personalized path planning in intelligent education system and will promote the research and application of adaptive learning technology.

AUTHOR CONTRIBUTIONS

Jinli Xuan designed, collected and analyzed the data, and drafted the manuscript. Jinli Xuan conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Jinli Xuan 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 received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] L. Elshani, "Personalized learning path generation based on genetic algorithms," UBT-ETD, 2021, DOI: 10.33107/UBT-ETD.2021.2649.
[2] F. Li and J. Sun, "Personalized learning path generation algorithm based on graph neural networks," Lecture Notes on Data Engineering and Communications Technologies, pp. 281-291, 2024, DOI: 10.1007/978-3-031-71619-5_24.
[3] D. Hauk and N. Soujon, "How reliable are large language models in analyzing the quality of written lesson plans? A mixed-methods study from a teacher internship program," Computers and Education: Artificial Intelligence, vol. 10, no. c, Art. no. 100538, 2026, DOI: 10.1016/j.caeai.2025.100538.
[4] Z. He, W. Xia, K. Dong, et al., "Unsupervised learning style classification for learning path generation in online education platforms," Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, DOI: 10.1145/3534678.3539107.
[5] X. Wei, W. Han, S. Sun, et al., "Educational psychology-empowered personalized learning path generation strategy," Learning Technologies, IEEE Transactions on, vol. 18, no. 000, pp. 741-756, 2025, DOI: 10.1109/TLT.2025.3590602.
[6] Y. Qu, L. Wang, H. Zhang, et al., "Personalized learning path generation based on neuro-cognitive collaborative filtering," 2024, DOI: 10.21203/rs.3.rs-4072040/v1.
[7] S. Ying and Y. Fu, "Knowledge graph construction and personalised learning path generation mechanism for online english teaching and learning," Applied Mathematics and Nonlinear Sciences, vol. 9, no. 1, 2024, DOI: 10.2478/amns-2024-3609.
[8] Y. S. Soekamto, L. C. Limanjaya, Y. K. Purwanto, et al., "From queries to courses: SKYRAG's revolution in learning path generation via keyword-based document retrieval," IEEE Access, vol. 13, pp. 21434-21455, 2025, DOI: 10.1109/access.2025.3535618.

[9] R. Zhang, X. Wu, X. Zhang, et al., “An Instructional Optimization Method Based on Bidirectional Transformer and Reinforcement Learning,” IEEE Access, vol. 13, 2025, DOI: 10.1109/ACCESS.2025.3576878.

[10] C. Tong and C. Ren, “Deep knowledge tracing and cognitive load estimation for personalized learning path generation using neural network architecture,” Scientific Reports, vol. 15, no. 1, 2025, DOI: 10.1038/s41598-025-10497-x.