Journal of Electronic Research and Application, 2025, Volume 9, Issue 3
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

An Analysis of the Construction Methods of Multimodal Course Knowledge Graphs

Fulin Li*

Guangdong University of Science and Technology, Dongguan 523000, Guangdong, China

**Author to whom correspondence should be addressed.*

Abstract: In the context of digitalization, course resources exhibit multimodal characteristics, covering various forms such as text, images, and videos. Course knowledge and learning resources are becoming increasingly diverse, providing favorable conditions for students' in-depth and efficient learning. Against this backdrop, how to scientifically apply emerging technologies to automatically collect, process, and integrate digital learning resources such as voices, videos, and courseware texts, and better innovate the organization and presentation forms of course knowledge has become an important development direction for "artificial intelligence + education." This article elaborates on the elements and characteristics of knowledge graphs, analyzes the construction steps of knowledge graphs, and explores the construction methods of multimodal course knowledge graphs from aspects such as dataset collection, course knowledge ontology identification, knowledge discovery, and association, providing references for the intelligent application of online open courses.

Keywords: Multimodality; Course knowledge graph; Construction method

Online publication: 5 June, 2025

1. Introduction

In the wave of educational digital transformation, the demand for knowledge management has become increasingly prominent. With the rapid development of information technology, course resources show multimodal features, including various forms such as text, images, and videos. Although this has greatly enriched the teaching content, it also brings many technical challenges. The heterogeneity of multimodal course resources makes data integration and management complex. The semantic relationships between different-modality data are difficult to measure directly, resulting in low-efficiency information retrieval and limited effective use of knowledge. At the same time, the large amount of data is sparsely distributed, and manual annotation is costly and prone to data loss. To address these challenges, it is particularly necessary to construct multimodal course knowledge graphs.

2. Overview of knowledge graphs

2.1. Core elements of knowledge graphs

The core of a knowledge graph lies in its unique triple-structure, namely entity-relationship-attribute [1]. Entities are the basic objects in a knowledge graph, representing specific things in the real world, such as concepts, people, and events in courses. Relationships describe the connections between entities, such as the sequential and causal relationships of knowledge points in courses. Attributes provide detailed characteristic information for entities, such as the difficulty level and credit hours of a course. This triple-structure enables knowledge to be represented and stored in a clear and structured manner, facilitating subsequent querying and reasoning.

2.2. Characteristics of multimodal knowledge graphs

Multimodal knowledge graphs are characterized by their fusion mechanism for heterogeneous data such as text, images, and videos. These different-modality data have unique features and expressions, and it is not easy to integrate them organically. Through advanced technical means, multimodal knowledge graphs can integrate the semantic information in text, the visual features in images, and the dynamic content in videos to form a more comprehensive and rich knowledge representation set [2].

3. Steps for constructing knowledge graphs

3.1. Ontology modeling

Ontology modeling is the key to constructing a knowledge graph. Subject experts need to form a modeling team. With their profound professional knowledge and rich teaching experience, they can accurately grasp the core and context of the course knowledge system. Through in-depth communication, the team can summarize and generalize the overall framework of the course, the distribution of knowledge points, and the logical relationships between knowledge [3]. On this basis, the team can design the concept layer, determine the core concepts, sub-concepts in the course, and their hierarchical structure; construct the data layer, map the design of the concept layer to the actual course data, ensure the specific data instances corresponding to each concept, as well as the attributes and relationships of these instances [4]. The mapping between the concept layer and the data layer needs to follow four principles. It is necessary to ensure integrity and accuracy, so that each data instance can be accurately mapped to the corresponding concept; at the same time, it is necessary to consider the diversity and flexibility of data to adapt to the characteristics of different course contents. The course knowledge system has distinct subject characteristics and involves many aspects, such as teaching objectives and teaching methods. Therefore, in the ontology-modeling process, it is necessary to fully consider the actual teaching needs, integrate the teaching objectives into the structure of the knowledge graph, and construct a knowledge graph that meets the course characteristics and teaching requirements [5].

3.2. Knowledge extraction from multisource data

When constructing a multimodal course knowledge graph, knowledge extraction from multisource data is a crucial link, especially the extraction strategies for structured teaching plans and unstructured videos [6]. Structured teaching plans usually have a clear structure and logic, including information such as course objectives, knowledge points, and teaching steps. For such data, traditional natural language processing (NLP) techniques can be used. Named entity recognition (NER) technology can be used to identify entities in the teaching plan, such as course names, knowledge-point names, and teacher names; a relationship-extraction model can be used

to mine the relationships between entities, such as the sequential and causal relationships of knowledge points. Unstructured video data is more complex and diverse. Videos contain rich visual and audio information and need to be processed with the help of computer vision (CV) technology [7]. For example, object-detection technology can be used to identify objects in the video, such as teaching equipment and demonstration models; behavior-recognition technology can be used to analyze the behavioral actions of teachers and students, such as explaining, asking questions, and discussing. To improve the efficiency and accuracy of knowledge extraction, a semi-automated annotation process should be implemented [8]. First, manually annotate a part of the data to provide training samples for the model. Then, use machine-learning algorithms to train on the annotated data to obtain a knowledge-extraction model. Next, use the trained model to automatically annotate a large amount of unannotated data, and conduct manual review and correction of the annotation results to improve the efficiency of knowledge extraction.

3.3. Knowledge fusion and quality verification

When constructing a multimodal course knowledge graph, knowledge fusion and quality verification are key links to ensure the accuracy and consistency of the graph. Entity disambiguation and co-reference resolution techniques are particularly important [9]. Entity disambiguation aims to solve the ambiguity of entities with the same name in different-modality data. By combining multimodal information, such as relevant scenes in pictures and videos, and the context relationships in the knowledge graph, the specific meaning of the entity can be accurately determined. Co-reference resolution is to identify the same entity pointed to by different expressions, such as “computer” and “PC,” and uniformly associate them with the same entity node in the knowledge graph to avoid knowledge redundancy and inconsistency. At the same time, a data-quality evaluation index system should be established to regularly check and correct the fused knowledge to ensure the accuracy and integrity of the knowledge graph [10]. Different-modality data may have inconsistent, repeated, or missing information. To resolve these conflicts, it is necessary to comprehensively consider the data source, reliability, and context information. Rule-based methods can be used to develop clear data-fusion rules to screen and integrate conflicting data; machine-learning algorithms can also be used to train a model to automatically handle data conflicts [11].

4. Construction methods of knowledge graphs for multimodal course content

4.1. Construction of multimodal datasets

The construction of multimodal datasets is the foundation for building multimodal course knowledge graphs. First-class course resources should be selected as the data-collection source. First-class course resources should be authoritative, systematic, and cutting-edge. The course knowledge should reflect the latest research results and development trends of the discipline, enabling learners to access the most advanced knowledge. During the collection process, attention should be paid to the diversity of resources, covering different types of courses, such as theoretical courses, practical courses, and experimental courses, as well as courses at different levels, such as basic courses, professional courses, and extended courses. There is a close relationship between video-frame extraction and text-corpus cleaning. Video-frame extraction is to extract key frames from course videos. These key frames contain important information in the videos, such as the teacher’s explanation content and the experimental process demonstrated. By analyzing the key frames, visual entities and relationships in the videos can be extracted, providing important data support for the construction of the knowledge graph. Text-corpus cleaning is to pre-

process the course text, removing noisy information such as typos, grammar errors, and redundant characters to improve the quality of the text. Establishing a metadata specification for course resources is an important part of constructing multimodal datasets. Metadata is data that describes data. It can provide detailed information about course resources, such as course names, instructors, course introductions, learning objectives, and target audiences. By establishing a unified metadata specification, standardized management of course resources can be achieved, facilitating resource retrieval and sharing [12].

4.2. Cross-modal ontology recognition technology

In the construction of multimodal course knowledge graphs, cross-modal ontology recognition technology is the key to realizing the fusion of different-modality data. It involves video-key-frame clustering algorithms, text-entity extraction and formula recognition, and the mapping mechanism between visual semantics and text semantics. Video-key-frame clustering algorithms are an important means for processing course-video data. Course videos usually contain a large number of frames, and directly processing these frames will bring a huge computational burden. Key-frame clustering algorithms aim to extract representative key frames from videos and cluster them to reduce the amount of data and discover the main content of the videos. A common method is clustering based on visual features. For example, features such as color, texture, and shape of key frames are extracted, and then a clustering algorithm (such as K-means clustering) is used to group similar key frames. In this way, the video content can be divided into different theme segments, and each segment corresponds to a cluster center, representing the main visual information of that segment. Text-entity extraction and formula recognition are two important tasks in processing course-text data, and their technical paths are different. Text-entity extraction mainly focuses on identifying entities with specific meanings from text, such as concepts, people, and events in courses. Common methods include rule-based methods, machine-learning-based methods, and deep-learning-based methods. Rule-based methods identify entities by manually writing rules, but this method has poor scalability; machine-learning-based methods require a large amount of annotated data for training, such as the conditional random field (CRF) model; deep-learning-based methods, such as pre-trained models like BERT, can automatically learn the semantic features of text and achieve good results in entity-extraction tasks [13]. Formula recognition focuses on identifying mathematical formulas from text and converting them into machine-understandable forms. Formula recognition usually requires the combination of optical character recognition (OCR) technology and formula-parsing algorithms. First, OCR technology is used to convert the formula images in the text into character sequences, and then the formula-parsing algorithm is used to analyze the character sequences to identify the structure and semantics of the formulas. To achieve the fusion of different-modality data, a mapping mechanism between visual semantics and text semantics needs to be established. Visual semantics mainly come from the results of video-key-frame clustering, while text semantics come from the results of text-entity extraction and formula recognition. A feasible method is to establish an intermediate semantic representation to map visual semantics and text semantics into the same semantic space. Through this mapping mechanism, different-modality data can be associated, providing a basis for the construction of multimodal course knowledge graphs.

4.3. Multidimensional knowledge association strategies

In the construction of multimodal course knowledge graphs, multidimensional knowledge-association strategies are the key to realizing the in-depth fusion and efficient use of knowledge, especially reflected in the course-

knowledge linking focusing on the spatio-temporal dimension, the dynamic matching between video-behavior recognition and text knowledge points, and the construction of a cross-modal inference rule base [14]. Course-knowledge linking focusing on the spatio-temporal dimension can provide learners with a more comprehensive and coherent knowledge system. In the time dimension, course content usually has a certain sequence and logical relationship. For example, basic concepts are explained first, followed by case analysis and practical operations. By establishing time-dimension knowledge links, knowledge points at different time points can be associated, helping learners better understand the development context and evolution process of knowledge. In the space dimension, different knowledge points in a course may be distributed in different textbook chapters, video segments, or pictures. Through space-dimension knowledge links, these scattered knowledge points can be integrated to form a complete knowledge network. The dynamic matching between video-behavior recognition and text knowledge points is an important means of realizing multimodal knowledge association. Course videos contain rich behavioral information, such as the teacher's teaching actions, demonstration operations, and students' interaction behaviors. Through video-behavior recognition technology, this behavior information can be dynamically matched with text knowledge points. The construction of a cross-modal inference rule base is the core of realizing multidimensional knowledge association. The cross-modal inference rule base contains the association rules and inference mechanisms between different modalities of data. Through these rules and mechanisms, inference and conversion from one modality of data to another can be achieved. At the same time, the cross-modal inference rule base can also support the expansion and update of knowledge. When new course resources are added, the system can automatically perform knowledge association and integration according to the rule base. By constructing a cross-modal inference rule base, the intelligence and automation of multimodal course knowledge graphs can be realized, providing more efficient and personalized learning services for learners.

5. Practical applications

5.1. Application in intelligent retrieval systems

The semantic-retrieval architecture based on knowledge graphs is the core of intelligent retrieval systems. This architecture is based on multimodal course knowledge graphs. Through the in-depth mining of entities, relationships, and attributes in the knowledge graph, it can achieve semantic understanding and matching of user queries. When a user enters a query, the system first performs semantic analysis on the query statement, extracts the key entities and relationships, then searches in the knowledge graph to find the matching knowledge nodes and associated relationships, and finally presents the retrieval results to the user.

5.2. Application in personalized learning services

Personalized learning-path planning relies on the association model between learner profiles and knowledge graphs to tailor-make learning plans for learners [15]. Learner profiles are constructed by collecting and analyzing data in many aspects, such as learners' learning behaviors, interest preferences, and knowledge levels. They can accurately describe the unique characteristics of each learner. Knowledge graphs provide rich knowledge resources and a clear knowledge structure for learning-path planning. By associating learner profiles with knowledge graphs, suitable learning content can be selected from the knowledge graph according to the specific situation of learners, meeting the personalized and diversified learning needs of students, and improving learning efficiency and effectiveness.

6. Conclusion

In conclusion, multimodal course knowledge graphs show significant value in the education field. They integrate multisource data, construct a structured knowledge system, provide strong support for intelligent retrieval and personalized learning, and improve the utilization efficiency of course resources and learning effects. In the future, the integration of 5G and VR technologies will bring innovative changes to knowledge services. The high-speed and stable 5G network ensures real-time data transmission, and VR technology creates an immersive learning environment, enabling learners to obtain knowledge immersively. The combination of multimodal knowledge graphs with these technologies is expected to provide more intelligent, efficient, and personalized educational services to teachers and students, promoting educational digitalization to a new level.

Funding

University-level Scientific Research Project in Natural Sciences "Research on the Retrieval Method of Multimodal First-Class Course Teaching Content Based on Knowledge Graph Collaboration" (GKY-2024KYYBK-31)

Disclosure statement

The author declares no conflict of interest.

References

[1] Feng Y, Liu X, Li K, et al., 2025, A Review of the Automatic Construction of Course Knowledge Graphs. Computer Technology and Development, 35(01): 1–11.

[2] Sun L, Meng F, Xu X, 2024, A Review of the Research on the Construction Technology of Course Knowledge Graphs. Computer Engineering, 2024: 1–25.

[3] Yan Y, Zhuang Y, 2024, Construction of Multimodal Knowledge Graphs and Their Applications in Intelligent Knowledge Services. Journal of Academic Library and Information Science, 42(06): 112–117.

[4] Guo Y, 2024, Research on the Construction Method of Course Multimodal Knowledge Graphs Based on Deep Learning, dissertation, Northeast Electric Power University.

[5] Zheng S, 2023, Course Knowledge Retrieval System based on Multimodal Knowledge Graphs, dissertation, Hebei University of Engineering.

[6] Zhang H, Song Y, 2023, Construction of Course Knowledge Graphs for Smart Education. Computer Education, (09): 120–125.

[7] Yan J, 2023, Research on Learning Path Recommendation of the U + Platform Based on Knowledge Graphs, dissertation, North University of China.

[8] Gao M, 2023, Research on the Construction and Application of Course Knowledge Graphs Integrating Multimodal Resources, dissertation, Inner Mongolia Normal University.

[9] Feng L, 2023, Design and Implementation of an Intelligent Teaching Aids Platform Based on Multimodal Knowledge Graphs, dissertation, Hubei Normal University.

[10] Zhao Y, Zhang L, Yan S, et al., 2023, A Review of the Construction and Application of Subject Knowledge Graphs in Personalized Learning. Computer Engineering and Applications, 59(10): 1–21.

[11] Liu H, Zhang G, 2022, Research on the Construction and Application of Course Knowledge Graphs. China

Educational Technology & Equipment, (22): 78–81.

[12] Gao M, Zhang L, 2022, Research on the Connotation, Technology, and Application of Educational Knowledge Graphs Integrating Multimodal Resources. Application Research of Computers, 39(08): 2257–2267.

[13] Wu H, 2021, Construction and Application of Personalized MOOC Courses Based on Multimodal Knowledge Graphs. University, (47): 64–66.

[14] Qi X, 2020, Research on the Construction and Application of Multimodal Course Knowledge Graphs, dissertation, Jilin University.

[15] Li Z, He F, Liu A, 2019, Construction and Application of Multimodal Teaching Knowledge Graphs. Fujian Computer, 35(08): 5–8.