Article

CourseKG: An Educational Knowledge Graph Based on Course Information for Precision Teaching

Ying Li [1], Yu Liang [2,*], Runze Yang [1], Jincheng Qiu [1], Chenlong Zhang [2] and Xiantao Zhang [1]

1   School of Computer Science and Engineering, Beihang University, Beijing 100191, China;
    liying@buaa.edu.cn (Y.L.); sy2206143@buaa.edu.cn (R.Y.); 17373255@buaa.edu.cn (J.Q.);
    21373285@buaa.edu.cn (X.Z.)
2   Faculty of Information, Beijing University of Technology, Beijing 100124, China;
    zhangchenlong@emails.bjut.edu.cn
*   Correspondence: yuliang@bjut.edu.cn

Abstract: With the rapid development of advanced technologies, such as artificial intelligence and deep learning, educational informatization has entered a new era. However, the explosion of information has brought numerous challenges. Knowledge graphs, as a crucial component of artificial intelligence, can contribute to the quality of teaching. This study proposes an educational knowledge graph based on course information named CourseKG for precision teaching. Precision teaching seeks to individualize the curriculum for each learner and optimize learning efficiency. CourseKG aims to establish a correct and comprehensive curriculum knowledge system and promote personalized learning paths. CourseKG can address the issue that current general-purpose knowledge graphs are not suitable for the education field. Particularly, this study proposes a framework for educational entity recognition based on the pre-trained BERT model. This framework captures relevant information in the educational domain using the BERT model and combines it with the BiGRU and multi-head self-attention mechanism to extract multi-scale and multi-level global dependency relationships. In addition, the CRF is used for character-label decoding. Further, a relationship extraction method based on the BERT model, which integrates sentence features and educational entities and estimates the similarity between knowledge pairs using cosine similarity, is proposed. The proposed CourseKG is verified by experiments using real-world C programming course data. The experimental results demonstrate the effectiveness of CourseKG. Finally, the results show that the proposed CourseKG can significantly enhance the precision teaching quality and realize multi-directional adaptation among teachers, courses, and students.

Keywords: knowledge graph; conception recognition; relation extraction; smart education

1. Introduction

With the development of artificial intelligence and the growing interest in the education field, smart education has attracted significant attention in recent years [1]. However, increasing digitization in the field of education has brought educators and institutions a multitude of new challenges [2,3]. Precision teaching driven by artificial intelligence, big data analysis, and other informational technology has become a development trend and a hot research topic in the educational information field. Precision teaching strives to optimize the efficiency of the learning process by customizing the curriculum for each learner [4,5].

The utilization of knowledge graphs (KGs) involves the reconstruction and reorganization of knowledge points derived from extensive course materials. The KG extracts and represents semantic issues of knowledge points based on the structure of nodes and edges in a KG [6]. In addition, the KG emphasizes knowledge organization, determining the optimal sequence for delivering knowledge throughout the curriculum. Using the KG, vast and fragmented knowledge can be seamlessly integrated to assist learners in intuitively

connecting within the knowledge system. This facilitates precision teaching, including the establishment of personalized learning paths.

Although KGs have demonstrated their efficiency in various domains of artificial intelligence, their application to the field of education has been relatively unexplored. Compared to general KGs, the construction of educational KGs faces many challenges. First, the expected educational concept entity is more abstract than the real-world entity. Second, the expected relationships are more cognitive and implicit and difficult to be deduced from the literal meaning of the text, which is the case with general KGs. In addition, domain adaptation in the education field has been hindered by the lack of relevant data, making it difficult to assess students' individual cognitive abilities. Further, the automation level in the KG construction process is relatively low and often relies on expert knowledge, which can lead to cognitive discrepancies among experts on the same topic, thus affecting scientific rigor and consistency.

Considering the above-mentioned challenges, this study proposes an educational KG based on course information, named CourseKG, for precision teaching. CourseKG employs advanced techniques, including deep learning and big data, for enhanced intelligence. It visually represents the points that should be taught in a lesson and their sequential relationships. Moreover, CourseKG can precisely assess a student's current learning level, determine their zone of proximal development, and identify their sweet spot.

With the KG acquired, CourseKG can attain precise teaching objectives by adhering to the principle of formulating teaching based on specific learning requirements. It enables the implementation of teaching tailored to students' aptitudes. The contributions of this study can be outlined as follows:

1. A precise teaching and education KG based on course information, named CourseKG, which uses heterogeneous course data, including structured, semi-structured, and unstructured teaching and learning assessment data, is proposed to extract teaching concepts and identify important educational relationships, effectively serving smart education.
2. An education entity recognition framework called the BERT-BiGRU-MHSA-CRF, based on a pre-training BERT model, is proposed. This framework uses the BERT model for word embedding in educational text and combines the BiGRU and multi-head self-attention (MHSA) mechanisms to extract global contextual relevance from multiple perspectives and levels. In addition, a CRF layer that considers the dependency relationships and constraints between character-level labels is used for decoding.
3. A relationship extraction method based on the BERT model is developed. This method combines sentence features and educational entities using the BERT model and estimates the similarity between knowledge pairs using cosine similarity.
4. Experiments conducted with real-world teaching data from C programming courses on two leading online learning platforms in China validate the scalability and feasibility of the proposed BERT-BiGRU-MHSA-CRF method. The experimental results demonstrate that the proposed method surpasses state-of-the-art approaches in both relation extraction and educational entity recognition.

The remainder of this paper is structured as follows: Section 2 provides an overview of the related work concerning the topic addressed in this study. Section 3 details the proposed methods and their implementation. Section 4 presents the experimental results and conducts an analysis. Finally, Section 5 summarizes the main conclusions drawn from the study and proposes directions for future research.

2. Related Work

2.1. General KG

In recent decades, KGs have become a hot research area since they can provide enterprises with structured data and factual knowledge, making their products more

intelligent [7]. In the development of KGs, the Semantic Web proposed by Berners-Lee [8] and Linked Data from Bizer [9] have made great contributions.

In 2012, the KG introduced by Google [10] improved the relevance of query results and users' search experience. Recently, numerous KGs have been developed, including the DBpedia [11,12], Wikidata [13], YAGO [14], and Freebase [15,16].

Since the introduction of knowledge graphs (KGs), much research has focused on information retrieval systems, recommendation systems, and question-answering (QA) systems [17]. QA systems based on KGs can be broadly categorized into three types: semantic-parsing-based systems [18], deep-learning-based systems [19], and embedding-based systems [20], along with information-retrieval-based systems [21]. It is worth noting that combining deep learning methods with traditional techniques can enhance the performance of KG-based QA systems [17]. KG-based recommendation systems typically fall into two categories: path-based methods [22] and embedding-based methods [23]. In recent years, information retrieval has garnered significant attention from various companies, such as Google and Facebook.

In addition to all the above-mentioned areas, KGs also play a key role in many other fields, including medicine, network security, finance, news, social networks, and education [17,24–29]. This work focuses on the educational KGs and their implementation in practice.

2.2. Educational KGs

In addition to the conventional fields of QA, RS, and IR, KGs have also achieved excellent results in the education field. Therefore, applying KGs and knowledge bases to the field of education has been very popular in recent research. Penghe proposed a system for the automatic construction of educational KGs [30]. In [31], the strength of semantic linking between knowledge points was assessed, and a Semantic Web-based learner preference model was proposed. In [32], the authors utilize machine learning to design a KG of courses, facilitating the study of MOOC courses. The relationship between the content expressed by KG technology was presented in a visual manner in [33]. Khan Academy [30] used educational KG for concept visualization and learning resource recommendation. In [34], a system for educational concepts extraction and identification of implicit relations focused on K-12 educational subjects was developed. A number of studies have suggested learning paths for learners based on a KG [35,36]. The main goal of [29] was to construct a knowledge map for the purpose of personalized university education. KGs can be applied to micro-learning [37]. Some researchers have constructed a knowledge base to assist in the decision-making process for the adaptation of micro open educational resources. KGs have been used in the teaching of courses, such as databases, as well as in other disciplines. However, the accuracy of educational KGs requires further improvement. In view of that, this work studies course information for achieving precision teaching.

Entity recognition is an essential step in constructing a KG, and extracting entities from unstructured or structured data based on predefined labels is the main task in this step [38]. In the beginning, entity recognition was mainly performed using methods based on rules and dictionaries and methods based on neural networks [39]. However, each of these two method types faces different problems. For instance, the former has low accuracy and poor portability, and the latter requires numerous manually labeled samples.

Relation extraction is another key part of the KG construction process; it is mainly performed to discern semantic connections between entities. Currently, the relation extraction methods mainly consist of unsupervised/semi-supervised, supervised learning, and template-based approaches [40]. Notably, supervised learning methods demonstrate superior performance. In addition, neural networks have been extensively employed in relation extraction to improve the performance of relation extraction tasks.

2.3. Current Problems

At present, the practical application of KGs to the field of education is still in its early stages, and there are several problems related to domain adaptation, knowledge granularity, and construction methods, which can be summarized as follows:

1. The knowledge category is coarse-grained: In typical KGs, nodes primarily represent real entities, leading to an uncertain granularity structure. Consequently, it becomes challenging to directly represent knowledge elements within a course.
2. Domain adaptation requires further improvement: The lack of an appropriate corpus in education poses a challenge to effectively simulate and test individual students' cognitive abilities at a granular level.
3. The degree of automation in constructing KGs is relatively limited: Numerous KGs heavily depend on expert knowledge. Addressing of the same knowledge point by various experts introduces cognitive variations, creating challenges in maintaining scientific rigor and consistency.

3. Proposed Method

In this study, a KG, which represents a type of semantic network, is used to extract and depict relationships among knowledge points within the course. This section describes in detail the construction flow of the proposed CourseKG, as shown in Figure 1.
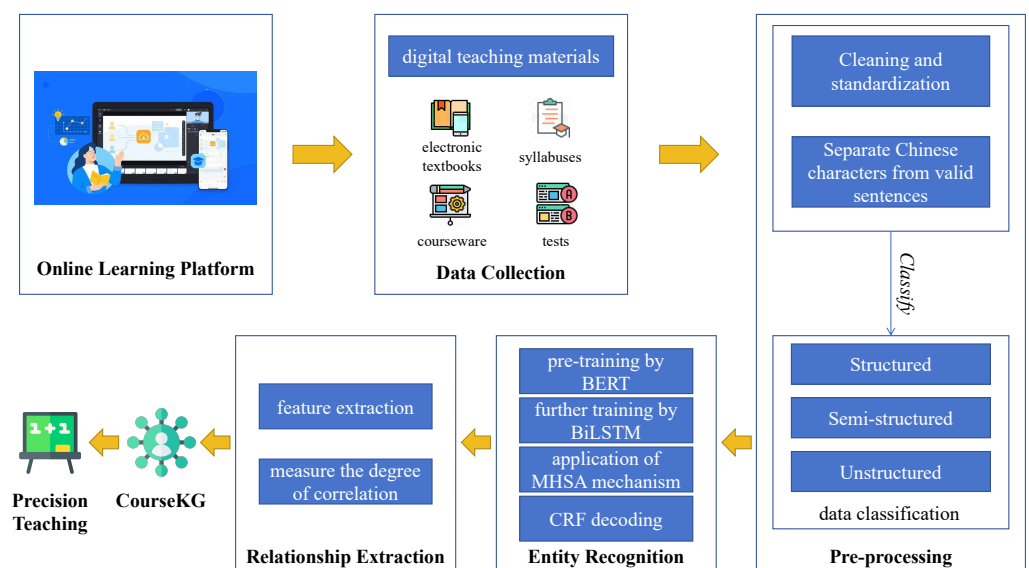
Figure 1. Block diagram of CourseKG.

We initially collected digital teaching materials from online learning platforms, which encompassed electronic textbooks, syllabuses, courseware, and tests. Subsequently, the pre-processing stage involved cleaning and standardization to separate Chinese characters from valid sentences. Entity recognition was then carried out through techniques such as BERT and BiLSTM, followed by feature extraction and measurement of correlation to extract relationships. Finally, the data were input into CourseKG for the purpose of precision teaching.

The process of converting raw data from an online learning platform to the high-quality CourseKG mainly includes three procedures: pre-processing, entity recognition, and relationship extraction. The three procedures are described in the following subsections.

3.1. Definition and Description

The definition of CourseKG is given by Definition 1.

Definition 1. *CourseKG* $=< Model, Data >$, *where Model* $=< V, P, R >$, *Data* $= d_a \cup d_b \cup d_c$

CourseKG aims to structure and integrate data based on the educational ontology type. Ontology typically includes a subclass-based taxonomic hierarchy, incorporating diverse classifications of concepts. The three elements of *Model*, *V* represents *a collection of extracted knowledge points*, *P* signifies *their properties or features*, and *R* denotes *relationships between entities*. The symbolic or numeric attributes of each point are depicted by nested feature vectors for establishing a standardized definition. The three elements are defined as follows:

- $V_i = [course, name]$, the entity name is defined based on educational ontology;
- $P_i = [type, objectives, difficulties]$, where:
  - $type = [knowledge, skills\_and\_abilities]$, denotes the nature of knowledge and can be categorized into three;
  - $objectives = [remember, understand, apply, analyze, evaluate, create]$; this is defined as a set of six hierarchical models following Bloom's taxonomy;
  - $difficulties = [easy, relativelyeasy, normal, hard, veryhard]$; this is employed to predict the level of difficulty for students in comprehending a specific knowledge unit V_i;
- $R_i = [inclusion, precursor, identity, brother, cause\_and\_effect, \dots]$. It is important to highlight that educational entities exhibit a sequential relationship and can possess numerous specific relations, as illustrated in Table 1.

Table 1. Special relations in *Model* of CourseKG.

Type	Description
inclusion	*A* contains *B*
precursor	*A* is a prerequisite for learning *B*
identity	*A* and *B* denote distinct descriptions of identical knowledge
brother	*A* and *B* share the same parent *C*, yet there is no sequential relationship between them
correlation	*A* and *B* are relevant but do not conform to the preceding relationships
inheritance	*A* is parent and *B* is its child
cause_and_effect	*A* triggers an event or condition, while *B* is the outcome arising from *A*

Further, *Data* denotes the foundation of CourseKG, sourced from three text types, as shown in Table 2.

**Table 2.** The three types of text in *Data* of CourseKG.

	Text Type	Description	Example
d_a	Structured	Contains data organized in a predefined format	*Multiple choice questionnaire, Table*
d_b	Semi-structured	Contains no structured data but has some structures	*Database, Text-survey, Syllabus*
d_c	Unstructured	Contains text-intensive data without a predefined data model	*Homework, Tutorial, Courseware*

Based on the above-presented data, *Model* of CourseKG has a hierarchical structure. One example of CourseKG is given in the following:

$$< V_i, P_i, R_i >=< [course.C\_programming, name.function],$$
$$[[type.skills, type.abilities],$$
$$[objectives.apply, objectives.analyze],$$
$$[difficulties.normal]],$$
$$[Precursor : \{function\_definition, function\_call, moduleprogramming\}],$$
$$[Brother : \{local\_variable, global\_variable\}] >$$

3.2. Data Pre-Processing and Knowledge Acquisition

Domain information embedded in the vocabulary can contribute to entity recognition contained in vocabulary and improve entity recognition performance. Given the scarcity of labeled data, the direct utilization of raw data for entity recognition in vertical domains for pre-training models like BERT [41] may not be as effective. To address this issue, this study first pre-processes raw data and performs knowledge acquisition. This can enhance the accuracy of subsequent entity recognition and contribute to more comprehensive knowledge mining and KG construction. In addition, based on the N-LTP [42], which is a language technology platform, the entity corpus construction process is performed, as shown in Figure 2.
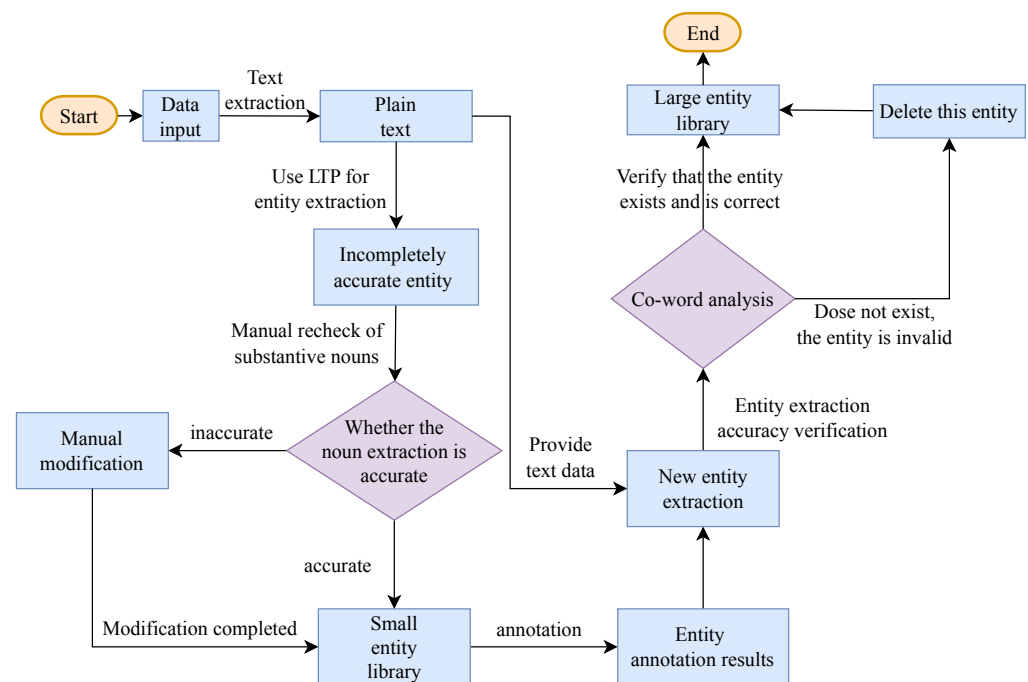

Figure 2. Flowchart of knowledge extraction from unstructured textual data.

3.3. Entity Recognition

This study proposes a framework called BERT-BiGRU-MHSA-CRF designed to enhance the accuracy of educational entity recognition within CourseKG. The structure of BERT-BiGRU-MHSA-CRF, as depicted in Figure 3, comprises four components: the BERT module, the BiGRU module, the MHSA module, and the CRF module. The BERT module is employed for word embedding of educational text. In this module, the BERT model converts Chinese characters into word vectors with textual information, achieving efficient embedding. The resulting embedded word vectors are input to the BiGRU module, where feature extraction is conducted on the word vectors. The MHSA module utilizes the MHSA mechanism to extract global feature correlation information from various perspectives

and levels. Finally, the CRF module fully considers intercharacter tag dependencies and constraints, decoding them with CRF to ensure the rationality of the final predicted tags.
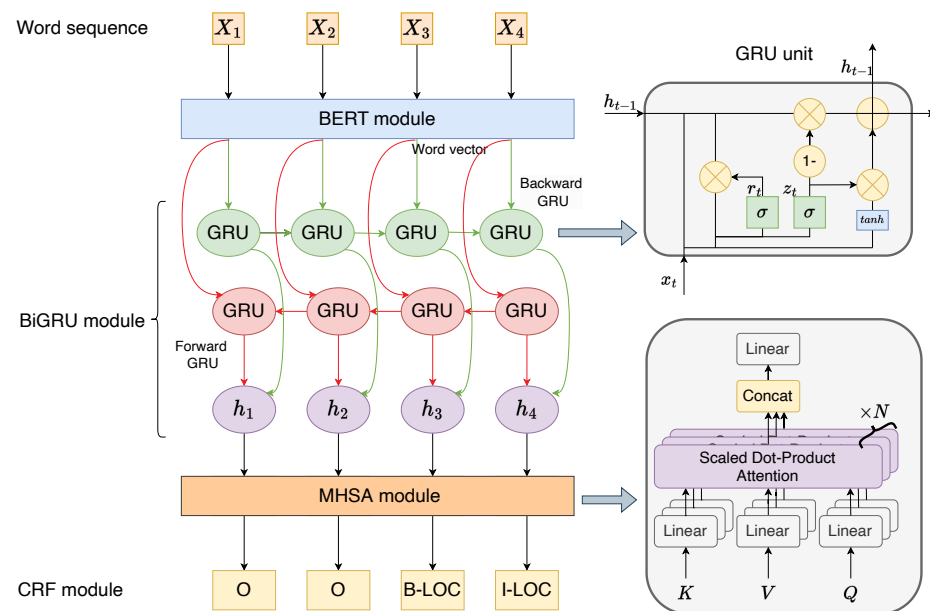
Figure 3. The entity recognition framework.

3.3.1. BERT Module

The BERT model is an excellent pre-training model for representing text word vectors. This model includes a multilayer bidirectional Transformer encoding that considers words before and after a particular word to determine its meaning in context. The BERT model shares a similar structure with the GPT and ELMO models. The Chinese BERT model is commonly acquired through unsupervised task training on extensive general-purpose corpora. Compared to the original BERT model, it acquires an enhanced feature representation of words and is applicable directly to downstream tasks.

In the field of education, there exists a widespread scarcity of corpora, and the available datasets fall short in supplying adequate data for BERT pre-training. Therefore, this study refines the BERT model through fine-tuning to enhance recognition accuracy. In addition, word embedding operations are performed on the training data to more effectively capture the content information embedded in educational texts. In the pre-training method used in this study, only the words related to education are masked. Finally, the word embedding vector is fed into the BERT model for feature extraction, yielding a sequence vector endowed with rich semantic features.

3.3.2. BiGRU Module

The gated recurrent unit (GRU) [43] is a variant of the long short-term memory neural network (LSTM). Traditional recurrent neural network (RNN) training frequently faces challenges such as gradient vanishing and explosion. In contrast, LSTM addresses these issues by incorporating input gates, output gates, and forget gates. Although the LSTM partially addresses the gradient disappearance problem, its computational process is time-intensive. The GRU structure streamlines the LSTM by merging the input gate and forget gate into an update gate. Consequently, the GRU preserves the strengths of the LSTM while simplifying its architecture. In the context of entity recognition in educational records, the GRU proves effective in extracting features.

Following the operating principle of the GRU unit, the GRU module excels at discarding redundant information, while its straightforward model structure reduces computational complexity. However, the naive GRU alone fails to fully leverage the contextual information present in educational records. Therefore, the backward GRU is introduced to

capture backward semantics in this paper. By incorporating both forward and backward GRU neural networks, the model, named the BiGRU model, aims to extract key features of named entities in educational records. The BiGRU model employs two sets of GRUs to extract features from sentences, ensuring that each token in a sequence is influenced by both its past and future contexts. This dual-directional approach provides access to both backward and forward information at each step.

The word vectors generated by BERT serve as the input to the BiGRU at each time step. At step t, a forward hidden layer processes the sequence from step 1 to step t, producing a forward hidden sequence $\overrightarrow{h_t} = \overrightarrow{GRU}$. Concurrently, a backward hidden layer handles the same sequence from t to 1, resulting in a backward hidden sequence $\overrightarrow{h_t} = \overrightarrow{GRU}$. Finally, the preceding and subsequent semantic features are combined to obtain $h_t = <\overrightarrow{h_t}, \overleftarrow{h_t}>$.

3.3.3. MHSA Module

The MHSA technique has been extensively applied in deep-learning-based natural language processing (NLP), especially in named-entity recognition tasks. When extracting text feature information, the integration of the MHSA mechanism proves effective in addressing the issue of time-series association in textual data. To enhance the extraction of interactive representation from a text sequence, the MHSA mechanism combines semantic feature information from various levels and perspectives, resulting in a comprehensive interactive representation of the text sequence.

In this study, the MHSA mechanism is focused on extracting feature information pivotal for entity recognition from the output of the BiGRU module. The matrix produced by the BiGRU comprises $Q_c, K_c, V_c \in \mathbb{R}^{1 \times B}$, where $c \in [1, \ldots, C-1, C]$. To establish correspondences between key slices and their respective query slices, the query–key pair computes their intermediate representational similarity using the scaled dot-product, followed by the SoftMax operation. Subsequently, the aggregation of self-attentional interactions is derived by multiplying the resulting matrix by V_c, yielding the representational context. For a single self-attention head h and a set of C slices, the representational context is determined as follows:

$$Head(Q_c, K_c, V_c)_h = Softmax\left(\frac{Q_c \cdot K_c^{Transpose}}{\sqrt{B}}\right) \cdot V_c \tag{1}$$

To compute the respective self-attention heads, multiple projections of Q_c, K_c, and V_c are employed, and their outputs are concatenated to achieve an aggregation of multiple heads, as depicted below:

$$MHSA(Q_c^{[1,\ldots,H]}, K_c^{[1,\ldots,H]}, V_c^{[1,\ldots,H]}) = Concat(Head_1, \ldots, Head_H) \tag{2}$$

The MHSA module consists of N MHSA layers. In the MHSA module, the long-distance characteristics and global information can be fully captured.

3.3.4. CRF Module

The BiGRU module performs well in tackling long-distance textual information. The CRF decoder [44] derives an optimal prediction sequence by leveraging the linking between adjacent entity labels. A significant strength of CRF is its capability to learn restrictive rules autonomously, thereby probabilistically reducing the occurrence of illogical sequences in the prediction sequence. In this study, the BIOES annotation method is adopted, where "B" signifies "beginning", "I" denotes "inside", "O" represents "out", "E" stands for "end", and "S" indicates "single". Each entity type is associated with its respective BIOES tags. The scoring function of the CRF decoder can be determined as follows:

$$score(X, y) = \sum_{i=1}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{i, y_i} \tag{3}$$

where X is the sequence $(x_1, x_2, ..., x_n)$; P and $A_{i,j}$ represent the observation and transition matrices, respectively; the scoring function sums the two matrices; and y corresponds to the label sequence of the predicted output.

The conditional probability $P(y|X)$ of y under a given X can be computed with scoring function as follows:

$$P(y|X) = \frac{\exp(score(X, y))}{\sum_{\widetilde{y} \in Y_X} \exp(X, \widetilde{y})} \tag{4}$$

where Y_X denotes all feasible label sequences for a given sentence.

During the decoding stage, the optimal sequence labeling y^* is determined by

$$y^* = argmax S_{\widetilde{y} \in Y_X}(X, \widetilde{y}) \tag{5}$$

3.4. Relationship Extraction

3.4.1. BERT-Based Method

The primary objective of this module is to discern the logical linkings within, thereby aiding learners in more effective learning. Relation extraction is crucial for learners: *inclusion*, *precursor*, *identity*, *brother*, *correlation*, *inheritance*, and *cause_and_effect* relationships. The relationship categories and their definitions are presented in Table 1. The locations of entities are very important in determining their relationships. According to the BERT model, sentence features and educational entities, previously identified in the preceding module, are fused.

For a given sentence $S = \{c_1, c_2, ..., c_n\}$, the head entity E_{head}, tail entity E_{tail}, and relationship R between them form a triple $< E_{head}, R, E_{tail} >$. As shown in Figure 4, the input includes a sentence and positions of the tail and head entities in the sentence, and the output is the relationship category. The word embedding vector sequence and sentence embedding vector are obtained from the last hidden state of the BERT, $M_{sentence}$. Then, the position information on the tail and head entities is used to encode their representation vector using the corresponding word embedding vector. Finally, the SoftMax layer performs classification prediction. The probability of the classification result for each category is decoded using the concatenation result of the head and tail entity vectors and the sentence vector as follows:
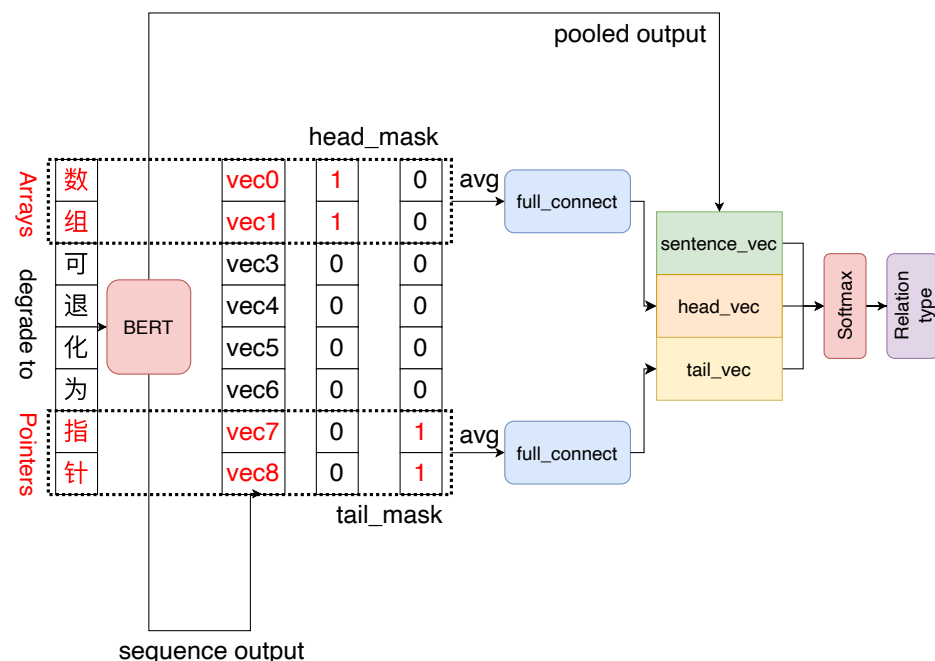

Figure 4. Block diagram of the relationship extraction method.

$$R = SoftMax(W \cdot \text{concat}(M_{sentence}, M_{head}, M_{tail}) + b), \tag{6}$$

where $R \in \mathbb{R}^{L*3d}$, L denotes total number of relationship types; d represents the hidden state size of the BERT model; M_{head} and M_{tail} denote hidden state vectors of the head entity E_{head} and tail entity E_{tail} in the last; and W and b are bias vectors.

3.4.2. Similarity-Based Method

Relation extraction addresses the challenge of semantic connection between entities. The predecessor–successor relationship stands out as a crucial link in knowledge, contributing to the generation of more coherent semantic sequences. Drawing from Mastery Learning [45], an educational philosophy and instructional strategy, asserts that students must achieve a certain level of mastery in prerequisite knowledge before progressing to learn subsequent information. When teaching parameters, arguments, and a function (knowledge A), it presupposes prior comprehension of the function (knowledge B). The theory suggests that one should master its predecessor knowledge A before understanding knowledge B. In this study, we use knowledge pairs to define knowledge points with sequential relationships. Numerous instances in KGs illustrate that the semantic relationship between a knowledge node and its neighboring nodes is often similar and strong. Consequently, this study employs pairwise similarity measurement [46] to extract their relationship. A cosine similarity approach is applied, as depicted in Equation (7). The larger the cosine value, indicating a smaller angle, the more related the two vectors are. The proposed CourseKG is interpreted as an undirected and weighted graph. Given the nodes V_i and V_j, $sim(V_i, V_j)$ denotes a weighted edge in CourseKG, where nodes V_i and V_j are connected if $sim(V_i, V_j) \neq 0$. The matrix $S_n = sim(V_i, V_j), (i, j = 1, \ldots, n-1, n)$ is referred to as the similarity matrix or affinity matrix, indicating pairwise similarities between knowledge points.

$$sim(\mathbf{A}, \mathbf{B}) = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2}\sqrt{\sum\limits_{i=1}^{n} B_i^2}} \tag{7}$$

4. Experiments

4.1. Data Collection and Pre-Processing

The experiment utilized data collected from two prominent online learning platforms in China, namely, CourseGrading [47] and Educoder [48]. Specifically, the experiments focused on a C programming course, involving the download of over 6000 teaching documents from the cloud platform. For experimental validation, we randomly chose 100 students as the subjects. The digital teaching material used to construct CourseKG comprised various text resources, including tests, syllabuses, electronic textbooks, and courseware.

Before training, we cleaned and formatted the textual data with a Chinese word segmentation tool, Jieba [49], ensuring compatibility with the Chinese BERT method. Finally, we extracted 6028 Chinese characters based on 27,042 valid sentences from the teaching materials. The raw and pre-processed data are shown in Table 3.

Table 3. The raw and pre-processed data. # stands for "number of".

Raw Data				Pre-Processed Data	
# electronic textbooks	# syllabuses	# courseware	# tests	# Chinese characters	# sentences
100	20	20	20	6028	27,042

For data pre-processing, the proposed BERT-BiGRU-MHSA-CRF framework was utilized to conduct BIOES annotation on a corpus consisting of 6028 Chinese words, thereby achieving knowledge point entity extraction. During model training, the process initiated with pre-training the model using the BERT module to obtain word vectors. Subsequently, these word vectors were fed into the BiLSTM module for additional training. Lastly,

the CRF decoding procedure was applied to predict the optimal label sequence for the BIOES module.

4.2. Entity Recognition Results

In the proposed entity recognition model BERT-BiGRU-MHSA-CRF, the model parameters were initially configured during the training phase. The batch size, which determines the size of the batch for each iteration, was set to 64, taking into account the dataset size and its influence on the gradient descent direction. To balance convergence speed and model fitting effectiveness, the learning rate was set to 0.0001, and the Adam optimizer was utilized. A dropout rate of 0.5 was employed. The model underwent training for 100 epochs, incorporating an early stopping strategy. The labeled datasets were partitioned into test and training sets at a ratio of 1:4. The parameters for the baseline models followed specifications from their original papers or initial implementations.

To evaluate the performance of the proposed BERT-BiGRU-MHSA-CRF, it was compared with two baseline methods: BERT-BiLSTM-CRF and BERT-GRU-CRF. The results of the proposed model and baselines are presented in Table 4.

Table 4. Comparison results of the three methods. The best results are emphasized in **bold**.

Method	Acc (%)	Precision (%)	Recall (%)	F1-Score (%)
BERT-BiLSTM-CRF	81.55	72.76	80.12	76.26
BERT-BiGRU-CRF	85.33	80.06	80.32	80.19
BERT-BiGRU-MHSA-CRF(ours)	**88.63**	**80.96**	**90.56**	**85.49**

Following the most recent studies, four evaluation metrics were selected to evaluate the models, namely, accuracy (*Acc*), recall, precision, and F1-score, and they were calculated by

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$F1\text{-}score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{11}$$

where TN represents the number of true negatives, TP denotes the number of true positives, FN corresponds to the number of false negatives, and FP signifies the number of false positives.

The experimental findings suggested that all models demonstrated effective performance on the dataset. The BiGRU method, which combines forward and backward GRU units to capture information both preceding and following a given sentence, exhibited enhanced utilization of contextual information. Specifically, the experimental results indicated superior performance of the BiGRU method compared to the BiLSTM method. Furthermore, the integration of a CRF layer facilitated automatic learning of sentence-level constraints and incorporation of constraint labels into the BiGRU output, thereby enhancing entity recognition performance. The accuracy and F1-score values achieved by the BERT-BiGRU-CRF model were 85.33% and 81.19%, respectively. Moreover, the addition of the MHSA module to the BERT-BiGRU-CRF model allowed for improved utilization of global information. Analysis revealed that the model augmented with the MHSA module exhibited increases in accuracy and F1-score values by 3.3% and 5.3%, respectively, underscoring the positive contribution of MHSA to named-entity recognition performance.

Owing to the absence of education-specific data, the conventional BERT model faced challenges in effectively extracting features from sequences. Education data had unique characteristics that could not be overlooked. This study used a fine-tuned BERT model, which enriched the contextual semantics in the vertical domain and provided domain

awareness, resulting in the best performance in educational entity recognition among all methods, having the accuracy and F1-score values of 88.36% and 85.49%, respectively.

Several measures were taken to prevent overfitting. First, we introduced regularization techniques to limit the complexity of model parameters and prevent overfitting to the training data. Secondly, we used an appropriate dropout value in the network to reduce the dependence between neurons and improve the generalization ability of the model. These comprehensive measures helped ensure that our models were less prone to overfitting and had better generalization performance when faced with new data.

The results suggest that the proposed BERT-BiGRU-MHSA-CRF method produced good outcomes. Moreover, this model was employed for entity prediction on unlabeled textual data, ensuring the quality of entity data and the KG.

4.2.1. Parametric Analysis

While training the model, there are two important parameters that need to be considered: the learning rate and the dropout value. If the learning rate is too large, then the model will converge too fast and may exceed the optimal value. If the learning rate is too small, then the model will converge too slowly and it may even cause the model to fail to converge. The dropout method can be used to avoid overfitting during model training. Based on the above considerations, this paper adds comparative learning rate and dropout value experiments to explore the model while also obtaining the best results.

First, by adjusting the learning rate continuously, the model effects were compared when the learning rate was 0.01, 0.001, and 0.0001. Table 4 shows the experimental results for different learning rates. For the model with a learning rate of 0.0001, the value of F1 was higher than that of the models with other learning rates, so the learning rate was chosen as 0.0001 based on the perspective of model performance. The dropout parameters for the experiment also need to be taken into consideration. In the forward propagation process, the dropout method causes a certain neuron to stop working temporarily according to a certain probability, P, which makes the generalization ability of the model stronger. Models can be regularized to some extent by not relying on local characteristics too much. The results of our model with different dropouts are shown in Table 5. The experimental results show that the model with dropout equal to 0.5 had better performance, so the dropout value of 0.5 was selected.

Table 5. Comparison results of different parameters. The best results are emphasized in **bold**.

Dropout	Learning Rate	Acc (%)	Precision (%)	Recall (%)	F1-Score (%)
	0.01	84.41	76.29	86.19	80.93
0.1	0.001	85.89	77.15	87.23	81.88
	0.0001	86.21	79.61	88.01	83.60
	0.01	85.12	77.34	89.30	82.89
0.5	0.001	86.43	78.79	90.08	84.06
	0.0001	**88.63**	**80.96**	**90.56**	**85.49**
	0.01	81.56	75.27	84.43	79.59
0.9	0.001	82.79	78.87	86.17	82.36
	0.0001	84.31	76.24	85.22	80.48

4.2.2. Efficiency Analysis

The accuracy and loss were also considered in the model training process. Figure 5 depicts the accuracy and loss curves, offering a visual representation of the model's performance throughout the training process. The accuracy curve shows the model's capability to accurately recognize data, whereas the loss curve reflects the incurred errors during training. Based on the results, both the loss and accuracy exhibit improvement as the number of epochs increases. This suggests that our model was progressively learning and refining its accuracy. After around 80 epochs, the model reached stable performance, indicating that additional training iterations may not yield substantial improvements. Consequently, the accuracy and loss curves illustrate that beyond approximately 80 training epochs, our model demonstrated strong convergence and excelled in performance.
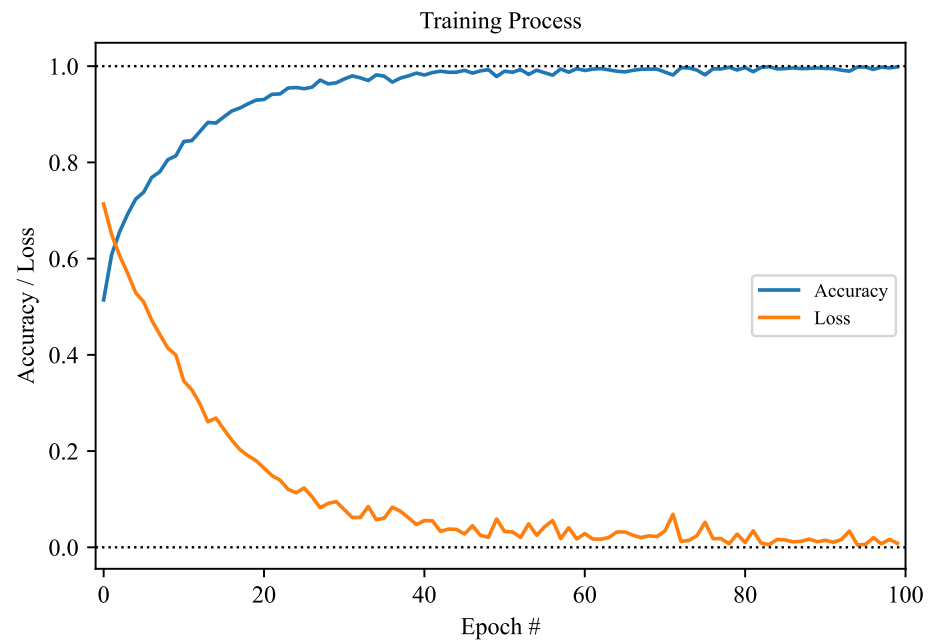
Figure 5. Accuracy and loss curves during the model training process.

4.2.3. Ablation Study

To examine the contributions of the three main modules in the proposed model, namely, the BERT, MHSA, and CRF modules, this study conducted an ablation study. The results obtained on the experimental dataset are presented in Table 6.

Table 6. The ablation study's results. The best results are emphasized in **bold**.

BERT	BiGRU	MHSA	CRF	Acc (%)	Precision (%)	Recall (%)	F1-Score (%)
✗	✓	✗	✗	66.39	54.97	50.34	52.56
✗	✓	✗	✓	77.65	69.69	69.99	69.84
✓	✓	✗	✓	85.33	80.06	80.32	80.19
✓	✓	✓	✓	**88.63**	**80.96**	**90.56**	**85.49**

The results for the CRF module show that the model performance could be improved by adding the CRF module to the model compared to employing the BiGRU method directly. The results indicate that the CRF module was beneficial for improving the entity recognition effect. Furthermore, a remarkable improvement in recognition performance is observed after introducing the BERT model fine-tuned with educational text information. This indicates that the educational semantic information derived from the BERT model could provide significant assistance in educational entity recognition. Moreover, the MHSA module comprehensively extracted global context information, leading to a certain improvement in the results.

*4.3. Extracted Relationship Results*

The connections among the knowledge points in CourseKG were established using the BERT model, and the similarity of all nodes was calculated. In this paper, a subset of the knowledge relationships of *Function* is depicted in Figure 6, where nodes represent the extracted knowledge point entity.
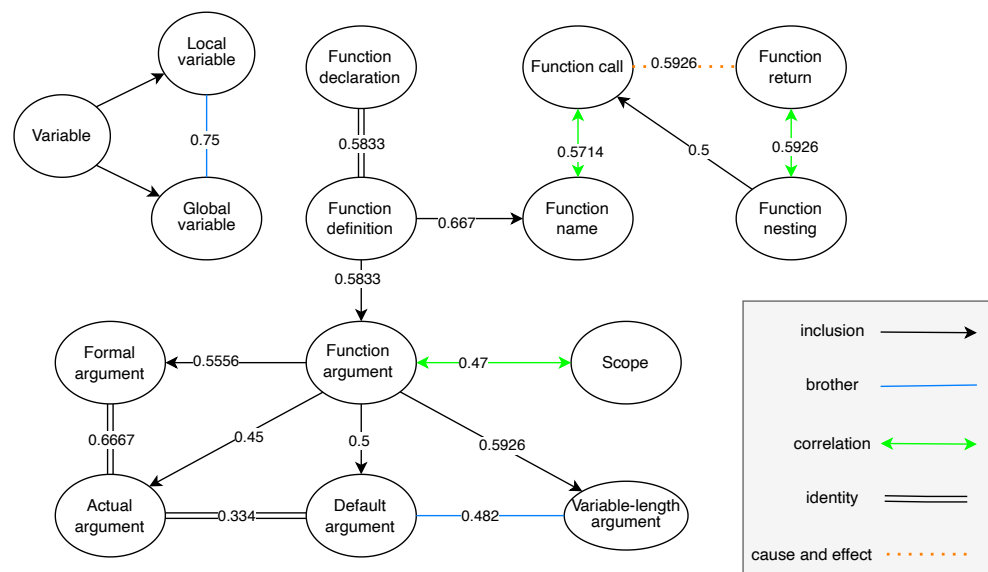
Figure 6. Extracted relationship results.

4.4. Visualization of CourseKG

In order to present CourseKG in a more intuitive and vivid manner, we employ Neo4j [50] as the backend graph database, and on the foundation of the Vue framework [51] integrate D3.js [52] to construct a visual representation of the knowledge graph, as illustrated in Figures 7–9.
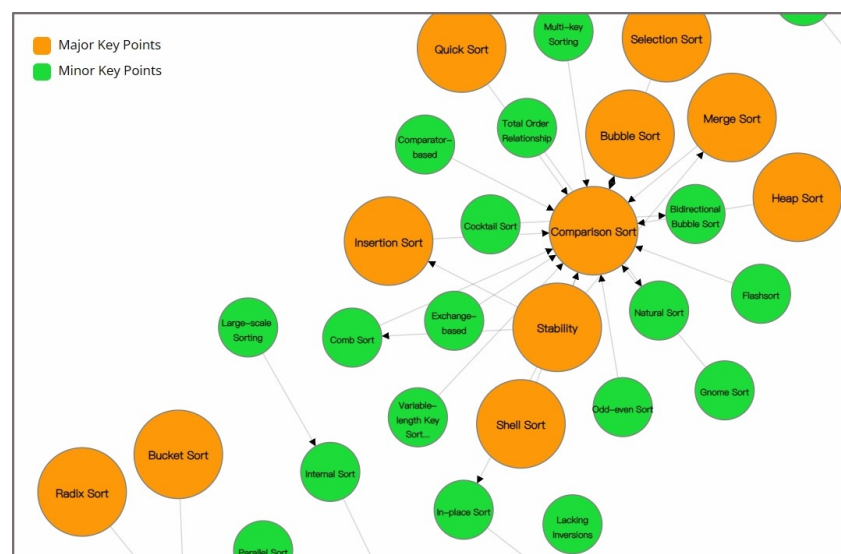

Figure 7. Visualization of knowledge graph. Orange denotes primary nodes, green represents secondary nodes.

This interactive visualization system allows users considerable freedom; they can drag, click, and access any relevant knowledge node at will. This knowledge graph is automatically constructed by our parsing pipeline, which has the capability to parse and analyze the desired textbooks.

To enable users to concentrate more effectively on the specific knowledge points of interest, we have implemented a focus differentiation feature with dynamic effects. When a user hovers their cursor over a particular knowledge point, that node and its directly associated knowledge points, along with the edges connecting them, are highlighted, while the opacity of the remaining knowledge nodes and edges is reduced. This design facilitates

effortless recognition of the currently focused knowledge node by the user. Moreover, a concise tooltip summary is generated for the current node, thereby allowing users to swiftly grasp the essence of the highlighted knowledge point.
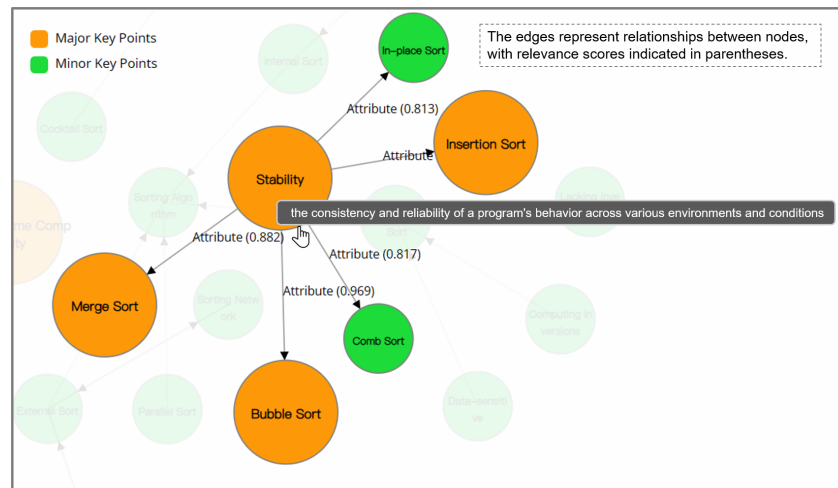
Figure 8. The focus differentiation in CourseKG. The cursor is currently hovering over the "Sorting" knowledge point, causing it and the directly interconnected knowledge points to be highlighted, with the opacity of other knowledge points diminished, thus facilitating clear distinction among them.

Furthermore, we also provide more detailed information for users who wish to delve further into the knowledge graph. CourseKG is capable of delivering a "relevance score" between any two nodes within a range of 0 to 1, where higher values denote a stronger correlation between the two knowledge points. This type of granular data serves as an expandable option, empowering users to gain a deeper understanding of the underlying knowledge system at hand.
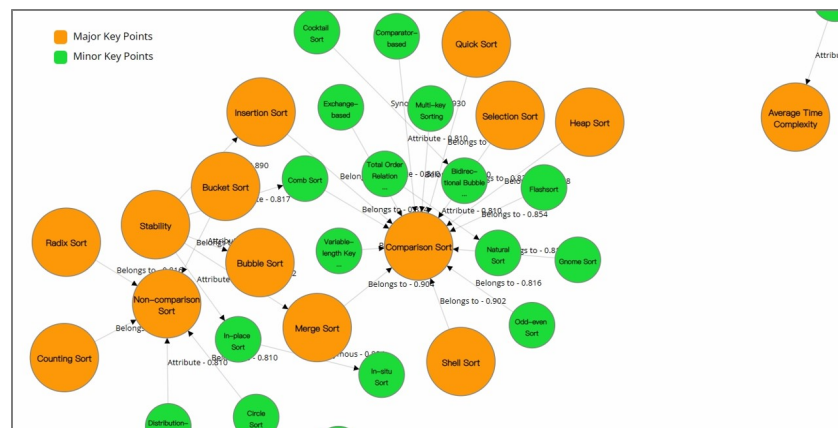
Figure 9. A visually augmented CourseKG, incorporating the representation of relationship associations. The scores between nodes are explicitly annotated on the edges, thereby enabling intuitive visualization and examination.

In essence, the provision of such scores allows for a nuanced exploration of the relationships among various knowledge points in the network, thereby enriching the user's investigative experience and enhancing their comprehension of the interconnectedness within the course's domain.

5. Conclusions and Future Work

This paper proposes CourseKG, utilizing diverse course data, enhancing smart education by extracting teaching concepts and identifying vital educational relationships.

We propose the BERT-BiGRU-MHSA-CRF framework for precise education entity recognition, integrating advanced mechanisms for global contextual relevance. Additionally, we present a groundbreaking relationship extraction method based on the BERT model. Experimental validation with real-world C programming course data demonstrates the scalability and superiority of our approach, surpassing state-of-the-art methods in both relation extraction and educational-entity recognition. Our research not only introduces innovative frameworks but also validates their effectiveness, emphasizing their potential impact on advancing knowledge representation in smart education.

Currently, our research primarily substantiates its experiments using a dataset from C programming courses. In the future, we can explore the possibility of collecting more teaching data across diverse academic disciplines and integrate various data sources to construct a more comprehensive knowledge graph. Beyond textual data, multi-modal data such as videos and audio can also furnish substantial informational support. Hence, future endeavors could investigate how to leverage multi-modal data to enrich the content of the knowledge graph and further enhance the efficacy of personalized learning.

We are fully aware that student feedback plays a pivotal role in personalized learning. Consequently, in the future, we may attempt to incorporate student feedback data and merge it with the knowledge graph to gain a deeper understanding of students' learning conditions and adaptively fine-tune instructional strategies accordingly. Moreover, natural language processing techniques can be employed to analyze student feedback, thereby extracting the most valuable insights for promoting individualized learning.

In order to improve the performance of CourseKG, future research can implement domain adaptation and transfer learning, combine domain knowledge, introduce dynamic context modeling technology, and explore continuous supervised learning methods. These innovative directions are expected to enhance the accuracy, robustness, and adaptability of the model so that it can better serve the diverse and evolving needs of the education field. The proposed CourseKG could offer educators a professional and efficient solution, enabling them to provide personalized and adaptive teaching experiences. Due to its precise instructional capabilities and personalized learning paths, the proposed CourseKG could revolutionize the delivery and reception of education.

Author Contributions: Conceptualization, Y.L. (Ying Li) and Y.L. (Yu Liang); methodology, R.Y.; software, J.Q.; validation, C.Z.; data curation, X.Z.; writing—original draft preparation, C.Z.; writing—review and editing, Y.L. (Yu Liang); visualization, X.Z.; supervision, Y.L. (Ying Li) and Y.L. (Yu Liang); project administration, Y.L. (Ying Li); funding acquisition, Y.L. (Ying Li). All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Public datasets in relevant fields were used and have been cited in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen, X.; Zou, D.; Xie, H.; Wang, F.L. Past, present, and future of smart learning: A topic-based bibliometric analysis. *Int. J. Educ. Technol. High. Educ.* **2021**, *18*, 1–29. [CrossRef]
2. Schmidt, J.T.; Tang, M. Digitalization in education: Challenges, trends and transformative potential. In *Führen und Managen in der Digitalen Transformation: Trends, Best Practices und Herausforderungen*; Springer: Wiesbadenb, Germany, 2020; pp. 287–312.
3. Anderson, T.; Rivera Vargas, P. A critical look at educational technology from a distance education perspective. *Digit. Educ. Rev.* **2020**, *37*, 208–229. [CrossRef]

4. Twyman, J.S. Personalizing learning through precision measurement. In *Handbook on Personalized Learning for States, Districts, and Schools*; Temple University: Philadelphia, PA, USA, 2016; pp. 145–164.

5. Li, Y.; Qiu, J.; Gui, S.; Song, Y.; Liu, T. Analytics 2.0 for Precision Education Driven by Knowledge Map. In Proceedings of the 2022 IEEE Frontiers in Education Conference (FIE), Uppsala, Sweden, 8–11 October 2022; IEEE: Toulouse, France, 2022; pp. 1–5.

6. Al-Moslmi, T.; Ocaña, M.G.; Opdahl, A.L.; Veres, C. Named entity extraction for knowledge graphs: A literature overview. *IEEE Access* **2020**, *8*, 32862–32881. [CrossRef]

7. Xiao, Y.; Zheng, S.; Shi, J.; Du, X.; Hong, J. Knowledge graph-based manufacturing process planning: A state-of-the-art review. *J. Manuf. Syst.* **2023**, *70*, 417–435. [CrossRef]

8. Berners-Lee, T.; Hendler, J.; Lassila, O. Web Semantic. *Sci. Am.* **2001**, *284*, 34–43. [CrossRef]

9. Bizer, C.; Heath, T.; Berners-Lee, T. Semantic services, interoperability and web applications: Emerging concepts. In *Chapter Linked Data: The Story So Far*; Information Science Reference: Hershey, PA, USA, 2011.

10. Singhal, A. Introducing the knowledge graph: Things, not strings. *Off. Google Blog* **2012**, *5*, 3.

11. Bizer, C.; Lehmann, J.; Kobilarov, G.; Auer, S.; Becker, C.; Cyganiak, R.; Hellmann, S. Dbpedia-a crystallization point for the web of data. *J. Web Semant.* **2009**, *7*, 154–165. [CrossRef]

12. Lehmann, J.; Isele, R.; Jakob, M.; Jentzsch, A.; Kontokostas, D.; Mendes, P.N.; Hellmann, S.; Morsey, M.; Van Kleef, P.; Auer, S.; et al. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semant. Web* **2015**, *6*, 167–195. [CrossRef]

13. Erxleben, F.; Günther, M.; Krötzsch, M.; Mendez, J.; Vrandečić, D. Introducing wikidata to the linked data web. In Proceedings of the The Semantic Web–ISWC 2014: 13th International Semantic Web Conference, Riva del Garda, Italy, 19–23 October 2014; Proceedings, Part I 13; Springer: Berlin/Heidelberg, Germany, 2014; pp. 50–65.

14. Suchanek, F.M.; Kasneci, G.; Weikum, G. Yago: A large ontology from wikipedia and wordnet. *J. Web Semant.* **2008**, *6*, 203–217. [CrossRef]

15. Bollacker, K.; Cook, R.; Tufts, P. Freebase: A shared database of structured general human knowledge. In Proceedings of the AAAI, Vancouver, BC, Canada, 22–26 July 2007; Volume 7, pp. 1962–1963.

16. Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; Taylor, J. Freebase: A collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, Vancouver, BC, Canada, 9–12 June 2008; pp. 1247–1250.

17. Zou, X. A survey on application of knowledge graph. In Proceedings of the Journal of Physics: Conference Series, Singapore, 17–19 January 2020; IOP Publishing: Bristol, UK, 2020; Volume 1487, p. 012016.

18. Berant, J.; Chou, A.; Frostig, R.; Liang, P. Semantic parsing on freebase from question-answer pairs. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Washington, DC, USA, 18–21 October 2013; pp. 1533–1544.

19. Dong, L.; Wei, F.; Zhou, M.; Xu, K. Question answering over freebase with multi-column convolutional neural networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 260–269.

20. Bordes, A.; Chopra, S.; Weston, J. Question answering with subgraph embeddings. *arXiv* **2014**, arXiv:1406.3676.

21. Yao, X.; Van Durme, B. Information extraction over structured data: Question answering with freebase. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 22–27 June 2014; pp. 956–966.

22. Wang, X.; Wang, D.; Xu, C.; He, X.; Cao, Y.; Chua, T.S. Explainable reasoning over knowledge graphs for recommendation. *Proc. Aaai Conf. Artif. Intell.* **2019**, *33*, 5329–5336. [CrossRef]

23. Wang, H.; Zhang, F.; Xie, X.; Guo, M. DKN: Deep knowledge-aware network for news recommendation. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 1835–1844.

24. Chandak, P.; Huang, K.; Zitnik, M. Building a knowledge graph to enable precision medicine. *Sci. Data* **2023**, *10*, 67. [CrossRef]

25. Liu, K.; Wang, F.; Ding, Z.; Liang, S.; Yu, Z.; Zhou, Y. A review of knowledge graph application scenarios in cyber security. *arXiv* **2022**, arXiv:2204.04769.

26. Wang, S. On the Analysis of Large Integrated Knowledge Graphs for Economics, Banking and Finance. In Proceedings of the EDBT/ICDT Workshops, Edinburgm, UK, 29 March–1 April 2022; Volume 3135.

27. Opdahl, A.L.; Al-Moslmi, T.; Dang-Nguyen, D.T.; Gallofré Ocaña, M.; Tessem, B.; Veres, C. Semantic knowledge graphs for the news: A review. *ACM Comput. Surv.* **2022**, *55*, 1–38. [CrossRef]

28. He, Q.; Yang, J.; Shi, B. Constructing knowledge graph for social networks in a deep and holistic way. In Proceedings of the Companion Proceedings of the Web Conference 2020, Taipei, Taiwan, 20–24 April 2020; pp. 307–308.

29. Rizun, M. Knowledge graph application in education: A literature review. *Acta Univ. Lodziensis. Folia Oeconomica* **2019**, *3*, 7–19. [CrossRef]

30. Chen, P.; Lu, Y.; Zheng, V.W.; Chen, X.; Yang, B. Knowedu: A system to construct knowledge graph for education. *IEEE Access* **2018**, *6*, 31553–31563. [CrossRef]

31. Su, Y.; Zhang, Y. Automatic construction of subject knowledge graph based on educational big data. In Proceedings of the 2020 the 3rd International Conference on Big Data and Education, London, UK, 1–3 April 2020; pp. 30–36.

32. Dalipi, F.; Imran, A.S.; Kastrati, Z. MOOC dropout prediction using machine learning techniques: Review and research challenges. In Proceedings of the 2018 IEEE Global Engineering Education Conference (EDUCON), Santa Cruz de Tenerife, Spain, 17–20 April 2018; IEEE: Toulouse, France, 2018; pp. 1007–1014.

33. Sun, K.; Liu, Y.; Guo, Z.; Wang, C. EduVis: Visualization for education knowledge graph based on web data. In Proceedings of the 9th International Symposium on Visual Information Communication and Interaction, Dallas, TX, USA, 24–26 September 2016; pp. 138–139.

34. Chen, P.; Lu, Y.; Zheng, V.W.; Chen, X.; Li, X. An automatic knowledge graph construction system for K-12 education. In Proceedings of the Fifth Annual ACM Conference on Learning at Scale, London, UK, 26–28 June 2018; pp. 1–4.

35. Shi, D.; Wang, T.; Xing, H.; Xu, H. A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning. *Knowl.-Based Syst.* **2020**, *195*, 105618. [CrossRef]

36. Shi, D.; Zhou, J.; Wang, D.; Wu, X. Research Status, Hotspots, and Evolutionary Trends of Intelligent Education from the Perspective of Knowledge Graph. *Sustainability* **2022**, *14*, 10934. [CrossRef]

37. Ma, N.; Zhao, F.; Zhou, P.Q.; He, J.J.; Du, L. Knowledge map-based online micro-learning: Impacts on learning engagement, knowledge structure, and learning performance of in-service teachers. *Interact. Learn. Environ.* **2023**, *31*, 2751–2766. [CrossRef]

38. Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; Dyer, C. Neural architectures for named entity recognition. *arXiv* **2016**, arXiv:1603.01360.

39. Yadav, V.; Bethard, S. A survey on recent advances in named entity recognition from deep learning models. *arXiv* **2019**, arXiv:1910.11470.

40. Aydar, M.; Bozal, O.; Ozbay, F. Neural relation extraction: A survey. *arXiv* **2020**, arXiv:2007.04247.

41. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.

42. Che, W.; Feng, Y.; Qin, L.; Liu, T. N-LTP: An Open-source Neural Language Technology Platform for Chinese. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Punta Cana, Dominican Republic, 7–11 November 2021; pp. 42–49. [CrossRef]

43. Dey, R.; Salem, F.M. Gate-variants of gated recurrent unit (GRU) neural networks. In Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, USA, 6–9 August 2017; IEEE: Toulouse, France, 2017; pp. 1597–1600.

44. Sutton, C.; McCallum, A. An introduction to conditional random fields. *Found. Trends Mach. Learn.* **2012**, *4*, 267–373. [CrossRef]

45. Block, J.H.; Burns, R.B. Mastery learning. *Rev. Res. Educ.* **1976**, *4*, 3–49.

46. Zhu, G.; Iglesias, C.A. Computing semantic similarity of concepts in knowledge graphs. *IEEE Trans. Knowl. Data Eng.* **2016**, *29*, 72–85. [CrossRef]

47. CourseGrading. Available online: https://www.educg.net/ (accessed on 10 March 2024).

48. Educoder. Available online: https://www.educoder.net/ (accessed on 10 March 2024).

49. Jieba. Available online: https://github.com/fxsjy/jieba/ (accessed on 10 March 2024).

50. Neo4j. Available online: https://neo4j.com/ (accessed on 10 March 2024).

51. Vue. Available online: https://vuejs.org/ (accessed on 10 March 2024).

52. D3. Available online: https://d3js.org/ (accessed on 10 March 2024).