information

Article

MEduKG: A Deep-Learning-Based Approach for Multi-Modal Educational Knowledge Graph Construction

Nan Li [1], Qiang Shen [1], Rui Song [2], Yang Chi [2] and Hao Xu [1,2,*]

[1] College of Computer Science and Technology, Jilin University, Changchun 130012, China; linan19@mails.jlu.edu.cn (N.L.); shenqiang19@mails.jlu.edu.cn (Q.S.)
[2] School of Artificial Intelligence, Jilin University, Changchun 130012, China; songrui20@mails.jlu.edu.cn (R.S.); yangchi19@mails.jlu.edu.cn (Y.C.)
* Correspondence: xuhao@jlu.edu.cn

Abstract: The popularity of information technology has given rise to a growing interest in smart education and has provided the possibility of combining online and offline education. Knowledge graphs, an effective technology for knowledge representation and management, have been successfully utilized to manage massive educational resources. However, the existing research on constructing educational knowledge graphs ignores multiple modalities and their relationships, such as teacher speeches and their relationship with knowledge. To tackle this problem, we propose an automatic approach to construct multi-modal educational knowledge graphs that integrate speech as a modal resource to facilitate the reuse of educational resources. Specifically, we first propose a fine-tuned Bidirectional Encoder Representation from Transformers (BERT) model based on education lexicon, called EduBERT, which can adaptively capture effective information in the education field. We also add a Bidirectional Long Short-Term Memory-Conditional Random Field (BiLSTM-CRF) to effectively identify educational entities. Then, the locational information of the entity is incorporated into BERT to extract the educational relationship. In addition, to cover the shortage of traditional text-based knowledge graphs, we focus on collecting teacher speech to construct a multi-modal knowledge graph. We propose a speech-fusion method that links these data into the graph as a class of entities. The numeric results show that our proposed approach can manage and present various modes of educational resources and that it can provide better education services.

Keywords: educational knowledge graph; multi-modal data; concept recognition; relation extraction; teaching speech fusion

Citation: Li, N.; Shen, Q.; Song, R.; Chi, Y.; Xu, H. MEduKG: A Deep-Learning-Based Approach for Multi-Modal Educational Knowledge Graph Construction. *Information* **2022**, *13*, 91. https://doi.org/10.3390/info13020091

Academic Editor: Paul Buitelaar

Received: 16 December 2021
Accepted: 11 February 2022
Published: 15 February 2022

1. Introduction

With the development of artificial intelligence and people's increasing emphasis on education, smart education has been drawing more attention in recent decades [1]. In recent years, various novel teaching manners have been utilized in college classrooms that leverage multimedia techniques, including textbooks, courseware, video, and voice, among other forms, rather than traditional methods such as blackboard writing. In these innovated educational methodologies, text is no longer the main form of knowledge dissemination, and multi-modal data such as pictures and audio are more conducive to students' understanding of knowledge [2–4]. Therefore, we need more intelligent methods/systems, through which to store, manage, and apply these multi-modal data.

Knowledge graphs serve as an important method, through which data can be organized and managed that interlinks heterogeneous data from different domains [5]. In the education field, knowledge graphs are often used for teaching and learning in schools. However, these knowledge graphs are frequently constructed manually, consuming a lot of resources, and they cannot be extended to other entities and relationships. Researchers have begun to focus on the automatic construction of educational knowledge graphs. Recent research [6–8] used knowledge graphs for ontology construction and achieved some

success. Liu et al. predicted the potential relationship between the concept and the course by mapping an online course to the general space of the concept [9]. Chen et al. proposed a system to construct educational knowledge graphs for students [10].

In general, most of the previous research data come from online education resources that are not integrated with real classrooms. Traditional educational knowledge graphs only utilize text as the only organizational form, which is monotonous and incomplete for the presentation of concepts or entity information. Compared to text, the use of pictures, the teacher's voice, and other modes of information make it easier for students to be interested in and understand the information being given to them in class. Therefore, the construction of multi-modal education knowledge graphs is particularly necessary and meaningful.

To tackle the challenges above, we propose a method that is able to automatically construct knowledge graphs that integrate multi-modal teaching resources, such as teacher speech. Taking a data structure course as an example, we used our method to realize the automatic integration of multi-modal educational resources. To improve the professionalism and domain of the educational knowledge graph, we propose a new model for educational entity recognition called EduBERT-BiLSTM-CRF. First, we build an educational lexicon and feed this into the fine-tuned BERT, which allows the BERT model to adaptively learn specific knowledge from the education field. Then, we use BiLSTM to extract the contextual features of each word in the input sentences. A CRF layer is added to obtain the optimal prediction sequence needed to complete education concept recognition. In addition, we use the location information of the entity to construct more accurate semantic relationships for these educational concepts. Finally, as entities in the graph occur in speech data, we convert classroom speech into text through speech recognition technology and link it to the knowledge graph as an entity. We also conduct extensive experiments, and the results show the effectiveness of our method. In summary, the main contributions of this research are as follows:

1. We propose a model to automatically construct a multi-modal educational knowledge graph, and we provide a way for speech fusion to incorporate and refine the knowledge graph by treating speech as an entity;
2. We propose a lexicon-based BERT model for educational concept recognition by combining the BiLSTM-CRF model that can better identify educational concepts. For relation extraction, in order to better combine the domain information, we combine the location information of the entity with BERT to dig out the implicit relationships between these entities;
3. We take computer courses as an example to verify the scalability and feasibility of our work. In addition, the empirical results show that our proposed approach performs competitively better than the state-of-the-art models in entity recognition and in relation extraction.

The rest of this paper is organized as follows: Section 2 introduces related work on knowledge graph. Section 3 briefly shows the details, which describes how the multi-nodal knowledge graph was built. Section 4 presents the experimental results. We summarize this research and discuss the prospects for future research in Section 5.

2. Related Work

This section introduces recent research on knowledge graphs and briefly describes named entity recognition and relation extraction technologies.

2.1. Educational Knowledge Graph Construction

In essence, a knowledge graph is a semantic network and a graphic set of related knowledge that generally refers to a large-scale knowledge base. At present, knowledge graphs are generally divided into general domain knowledge graphs and vertical domain knowledge graphs. Examples of classic general domain knowledge graphs include YAGO [11], DBpedia [12], Wikidata [13], etc. These general domain graphs have great advantages in semantic search, question answering systems and other scenarios, however

they also have disadvantages. They cannot support the organization and management of entities in specific fields well, as they require deep domain knowledge. Vertical domain knowledge graphs play an important role in this respect, however vertical domain knowledge graphs are often manually constructed, requiring a lot of time and human resources [14].

Recently, knowledge graph technology has played an important role in the education field [15]. Yang et al. used the correlation between specific courses to establish a directed universal concept graph and to explore the implied correlation between courses [16]. Senthilkumar introduced a concept map constructed by software into teaching and learning [17]. Liang et al. explored the prerequisite relationships of concepts by mining dependencies between courses [18]. These studies have proved that it is very important to dig out educational concepts and relationships. Many researchers have begun to try to construct knowledge graphs by integrating a large amount of educational data. Su et al. constructed a subject knowledge graph that evaluates the strength of the semantic associations between knowledge points [15]. Sun et al. used sub-string matching for entity recognition and the clustering method for semantic relation extraction to build a visual analysis platform called EduVis [19]. Zheng et al. constructed a curriculum knowledge graph by using Vector Space Model(VSM) and rules processing to learn easily [20]. Dang et al. used Wikipedia for entity extraction and constructed an MOOC knowledge graph [21] Yao et al. proposed a novel model for embedding the learning of educational knowledge graphs to promote knowledge graph construction [22].

These studies have proved the urgency for the construction of knowledge graphs in the education field. However, existing works have only used educational data from a single mode, such as course outlines and other text resources. They ignore other modal educational data from offline real classrooms, such as teacher audio, meaning that it is possible that these knowledge graphs lack integral information. In addition, previous studies did not fully realize automatic knowledge graph construction. They required a lot of manual annotations and templates. Therefore, the goal of the present research is to design a multi-modal educational knowledge graph model that automatically combines online and offline real resources, which can effectively serve smart education.

2.2. Named Entity Recognition

Named entity recognition (NER), a key step in the construction of knowledge graphs, aims to extract entities from structured or unstructured data according to predefined tags [23]. Research on entity recognition in the vertical domain has been drawing more attention in recent decades.

Previously published work on named entity recognition is mainly divided into rule-based and dictionary-based methods, machine learning-based methods, and neural network-based methods [24]. The rule-based and dictionary-based approach was first applied to NER, which involves rules being manually written in order to identify entities by matching text to rules [25]. Manual writing requires a lot of time, low accuracy, and poor portability. Machine learning models can solve the above problems well. Common machine learning models include Hidden Markov Model (HMM) [26], Maximum Support Vector Machine (SVM) [27], Conditional Random Field (CRF) [28], etc. These methods require manual feature extraction. Model training requires a large number of manual labeling samples, and the results are not ideal. At present, entity recognition tasks usually use neural network models to build sequence labeling models and to automatically extract features. Classical encoders include Convolution Neural Networks (CNN) [29], Bidirectional Long Short-Term Memory (BiLSTM) [30], and their variants [31–34].

Recently, pre-trained language models (PLMs) have made historic breakthroughs in many natural language processing tasks, such as Embeddings from Language Models (ELMO) [35], Generative Pre-training (GPT) [36], and Bidirectional Encoder Representation from Transformer (BERT) [37]. It is undeniable that a pre-trained BERT model performs well in general domains, however is inefficient in specific domains. Considering the shortage

of data resources in the education field, fine-tuning can help the model learn the domain-related knowledge better. In addition, an adaptive embedding is constructed by lexical enhancement. We added a BiLSTM model to capture two-way semantic dependency. The CRF model can automatically learn constraint information through the training corpus and can avoid illegal sequences in the prediction results. In our paper, we chose CRF as the decoder to obtain educational entities.

2.3. Relation Extraction

Relation extraction is another key step in graph construction that aims to identify the semantic relationships between entities. Currently, there are several main types of methods: template-based methods, supervised learning methods, and semi-supervised/unsupervised methods [38]. Among these, the supervised relation extraction method has demonstrated the best performance.

Neural networks have been widely used in relation extraction, such as in CNNs [29], and in Recurrent Neural Networks (RNN) [39]. He et al. constructed a system with a novel deep neural network (DNN) to automatically infer associations in the biomedical-related literature [40]. Zeng et al. proposed a novel model dubbed Piecewise Convolutional Neural Networks (PCNNs) with multi-instance learning [41]. Zhang et al. used BiLSTM and the features derived from lexical resources [42]. A Graph Neural Network (GNN) [43] is a kind of neural network that can capture the topological characteristics of graph data. GNN-based models [44–46] usually use the text dependency trees as the input to an a-priori graph structure in order to obtain richer information expression. Using language models to better express the relationship, some work regards relation extraction as a downstream taskPLM. Wu et al. proposed a model that both leverages the BERT model and incorporates entity information to tackle relation classification tasks [47]. Cheng et al. designed a new network architecture with a special loss function as a downstream PLM model [48]. These language models have achieved good results in terms of relation extraction. Generally speaking, relation extraction depends on the information in the sentence and the target entity. Based on the idea of the R-BERT model [47], in this paper, we use an education-based lexicon to pre-process the corpus and introduce entity location information into the BERT model that can better fuse sentence and lexical features to achieve relation extraction tasks.

3. Methods

In this section, the detailed method is introduced. First, we introduce how to construct a multi-modal educational knowledge graph. Second, we describe the entity recognition and relation extraction methods in detail. Finally, we demonstrate how to utilize teacher speech to build the knowledge graph.

3.1. Framework Overview

In order to better construct a knowledge graph for a specific educational field, we divided our system framework into the following three modules: an educational concept recognition module, an educational relation extraction module, and a teacher speech-fusion module. A diagram of the framework is shown in Figure 1. The modules can be described as follows:

- Educational Concept Recognition Module: The main goal of this module is to extract teaching concepts or educational entities in a specified course. Online education resources include Baidu entries and Jianshu articles. Offline education resources usually include course outlines, PowerPoints, and teaching courses. The lexicon enhancement method is used to pre-process the data, and then combine a fine-tuned BERT model to extract the educational concepts. The final outputs of this module are the extracted concepts, which are the basis for the construction of the knowledge graph;
- Educational Relation Extraction Module: The main goal of this module is to associate the extracted educational concepts to help learners clarify the relationship between knowledge concepts. Vocabulary information is still important for relation classifica-

tion. This module uses the acquired entities for vocabulary labeling and combines itself with the BERT model to distinguish the potential relationships between educational concepts;

- Teacher Speech Fusion Module: The teacher's voice is also an important resource in the field of education. The main goal of this module is to fuse real classroom teacher speech as a kind of entity into the text-based education knowledge graph. The module mainly uses Mel Frequency Cepstral Coefficients (MFCC) to extract speech feature variables and performs the Fourier transform. HMM is used to obtain speech text and calculate similarity. Teacher speech is matched with text entities.

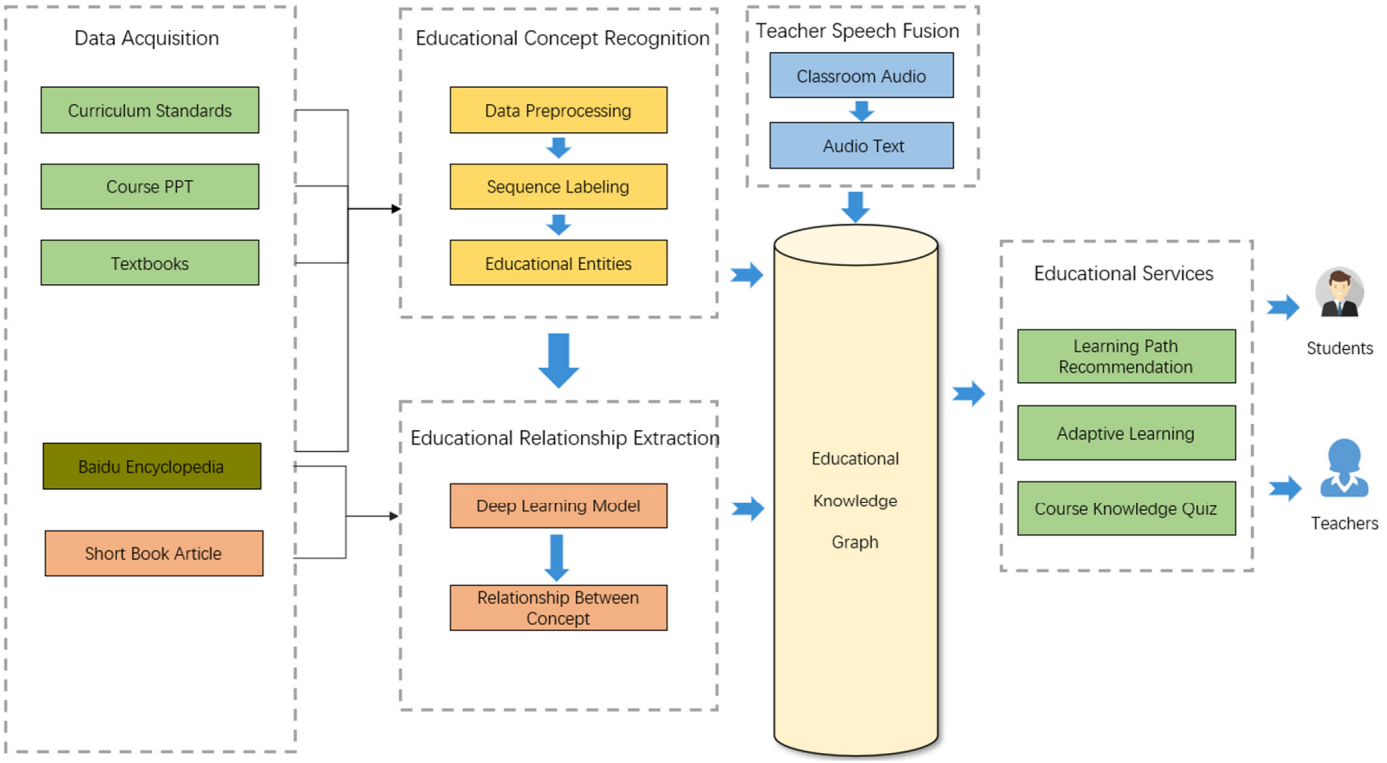


Figure 1. Construction framework of our multi-modal educational knowledge graph.

3.2. Educational Concepts Recognition

The domain information contained in vocabulary can help with entity recognition performance. Due to the shortage of labeling data resources, pre-training models such as BERT use them directly for entity recognition in the vertical domain, which may not be effective. First, we created an education vocabulary. Three domain experts annotated the entities in the course outline. When two or more experts marked the same entity type, we regarded it as the final type of the entity. The vocabulary consisted of these entities. Then, we proposed a model called EduBERT-BiLSTM-CRF to improve the accuracy of educational concept recognition by combining the educational lexicon to encode the characters as well as a fine-tuned BERT model. Figure 2 shows the architecture of our model, which consists of four parts. The first part is the character representation module; the second part is the fine-tuned BERT module; the third part is BiLSTM module; and the fourth part is CRF module. First, each character in a sentence corresponds to a dense vector. According to the domain lexicon that we built, all of the vocabulary information corresponding to each character is added to the representation of each character. Then, these enhanced characters are introduced into fine-tuned BERT, and word embeddings can be learned by BiLSTM. Finally, the output of the BiLSTM model is decoded to obtain the optimal label sequence.

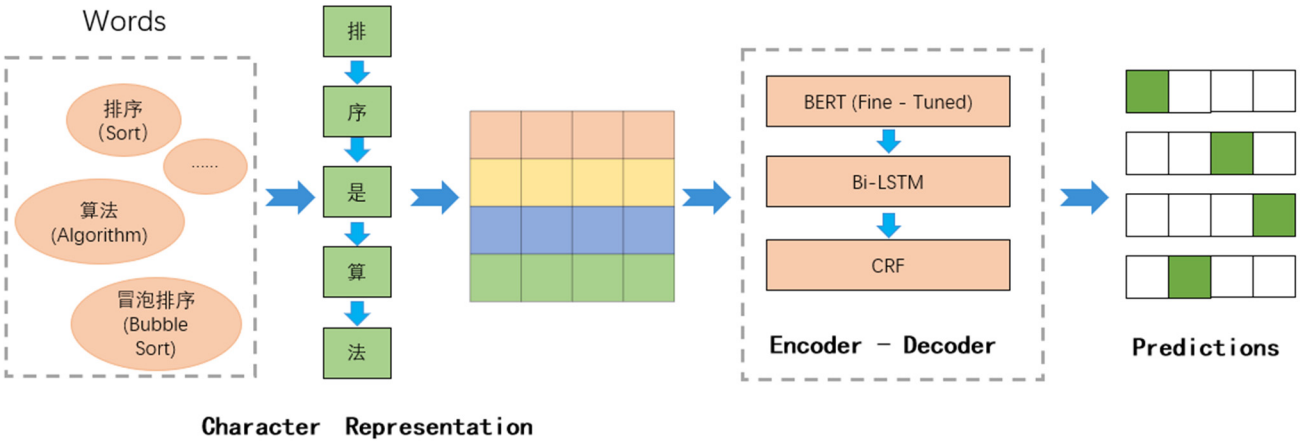


Figure 2. The architecture of the EduBERT-BiLSTM-CRF model for educational concept recognition. The input sentences are able to obtain four types of tag sets according to the information in the vocabulary. These vectors are used as the fine-tuned BERT input, and they are then encoded and decoded by BiLSTM and CRF to complete sequence annotation. The Chinese input is "排序是算法" (sorting is an algorithm).

3.2.1. Character Representation

Given the converted educational data, this concept extraction task can be viewed as a word sequence labeling problem. To preserve as much vocabulary as possible for all of the characters, we defined four word label sets: (1) *B*: "contains all words starting with this character"; (2) *I*: "contains all words with this character in the middle"; (3) *E*: "contains all words ending with this character"; and (4) *S*: "vocabulary consisting of only this character". Figure 3 shows an example.

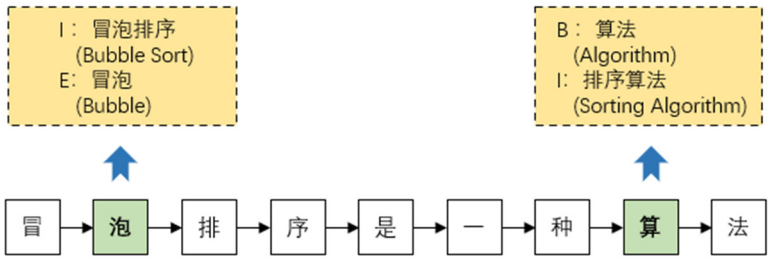


Figure 3. An example to illustrate tag collection. The character "泡" occurs in two words, and it occurs in the middle of "冒泡排序" and at the end of "冒泡". Therefore, the I-label set is "冒泡排序" and the E-label set is "冒泡". The Chinese input is "冒泡排序是一种算法" (bubble sort is an algorithm).

The input sequence is seen as a character sequence $S = \{x_1, x_2, \ldots, x_n\}$, and each character x_i $(1 \leq i \leq n)$ is a four-word set. To utilize the lexicon information effectively, the collected vocabulary is compressed using word weighting. Its equation is

$$v^s(S) = \frac{1}{Z} \sum_{w \in S} (z(w) + c) e^w(w) \tag{1}$$

$$Z = \sum_{w \in B \cup I \cup E \cup S} z(w) + c \tag{2}$$

Here, S denotes a word set, e^w denotes the word embedding, $z(w)$ denotes the word frequency, Z is the four-class label weight normalization, and c denotes the number of the word sets.

The data set consists of a training set and a test set. The frequency of the character will not increase if a short word composed of x_i is covered by a long word. This avoids the problem where the frequency of a short word is always less than the frequency of the long word covering it. We use "链表" (linked list) and "双向链表" (double-linked list) as examples. When calculating the word frequency of the double-linked list, the word frequency of the linked list does not increase as the linked list and double-linked list overlap. By embedding vocabulary collections into characters, the model can make better use of character information and vocabulary information.

3.2.2. Fine-Tuned BERT

To make better use of contextual semantic features, we use BERT as the generator to generate word embedding as input to the next module. Unlike previous pre-trained models, such as Word2vec [49], the BERT model combines the advantages of the ELMO [35] and GPT [36] models. Instead of using the traditional one-way language model or shallow splicing of two one-way language models for pre-training, it uses a multi-layer bidirectional transformer network structure to generate bidirectional semantic features. A Bidirectional transformer encoder is the key structure of BERT, and its main role is the self-attention mechanism. It computes the attention function for a set of queries and packs it into matrix Q. The keys and values are stored into matrices K and V. Self-attention adjusts the weight factor matrix to obtain the representation of words based on the degree of correlation between words in the same sentence:

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

where $Q = K = V$ and d_k is the embedding dimension.

The multi-head attention mechanism projects Q, K, and V through several different linear transformations, and finally, it stitches together different attention results. Its equations are as follows:

$$\text{MultiHead}\ (Q, K, V) = \text{Concat}\ (\text{head}_1, \cdots,\ \text{head}_n)W^O \tag{4}$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{5}$$

where W_i^Q, W_i^K, W_i^V, and W^O are parameter matrices.

The Chinese BERT model proposed by Google was trained using Chinese characters. These characters are randomly blocked when the model generates training samples. It does not consider traditional Chinese word segmentation, so this paper introduces a new mechanism: the whole word mask; that is, if a part of the word that belongs to the same word is blocked, then the other parts of the word are also blocked. Table 1 shows an example.

Table 1. An illustration of the current mask mechanism. The Chinese input is "排序是一种算法" (sorting is an algorithm). The result of the input sentence after word segmentation is several words: "排序" (sort), "是" (is), "一种" (a), and "算法" (algorithm). Using the original MASK mechanism, only part of the words can be masked, such as "排[M]" and "[M]法". Using the current MASK mechanism, the whole word can be masked. For example, "排序" can be masked by the token [M].

Input	排序是一种算法 **(Sorting Is an Algorithm)**
Word Segmentation	排序 是 一种 算法
Original Mask	排[M] 是 一种 [M]法
Current Mask	[M][M] 是 一种 [M][M]

For the education field, there is a general lack of corpus, and the existing corpus cannot provide enough data for BERT pre-training. Therefore, we used fine-tuned BERT to improve recognition accuracy. A fully connected network was used at the top of BERT, obtaining 768-dimensional context representation. The BERT model in this paper consists of 12 layers, each of which fuses the semantic information of the context. We fine-tuned the last four layers. We masked part of the word sequence as a whole word, and marked the beginning of the sentence with a [CLS], separating the sentence from the sentence with a [SEP]. The output embedding of each word consists of three parts: token embedding, segment embedding, and position embedding. These embeddings can make better use of lexical features and sentence features. Sequence vectors are inputted into the bidirectional transformer for feature extraction, and then sequence vectors with rich semantic features are obtained.

3.2.3. BiLSTM Encoder

BiLSTM is used to extract features from these sentences through two LSTMs, and for learning, each token in the sequence is based on both the future and the past context of the token. Both forward and backward information is available for each moment.

The BERT output word vector is used as the input for each BiLSTM time step. At each time step t, a hidden forward layer manages the sequence from step 1 to step t, obtaining a forward hidden sequence ($\overrightarrow{h}_1, \overrightarrow{h}_2, \overrightarrow{h}_3, \ldots \overrightarrow{h}_t$), and a hidden backward layer with the same sequence as step t to step 1, obtaining a backward hidden sequence ($\overleftarrow{h}_1, \overleftarrow{h}_2, \overleftarrow{h}_3, \ldots, \overleftarrow{h}_t$). Hidden layer state sequence stitching is generated, that is, $h_t = [\overrightarrow{h}_t : \overleftarrow{h}_t]$. By matrix changing the output sequence, the hidden state sequence is mapped into the k dimension (k is the number of categories of labels) via the linear output layer. The mapping matrix $Q = (q_1, q_2, \ldots, q_n) \in R^{n*k}$ are the combined outputs, where q_n and k are the scores of the k-th label relative to the n-th category.

3.2.4. CRF Decoder

The BiLSTM model is good at handling long-distance text information. The CRF model can use the relationship between adjacent entity labels to obtain the optimal prediction sequence. The biggest advantage of CRF is to reduce the probability of irrational sequences in the prediction sequence by automatically learning restrictive rules. In the paper, we use different tokens to represent three types of entities: "a" for algorithms, "s" for structures, and "c" for basic terminology. Each type of entity is used for its own BIEO tags, such as B-a, I-a, I-s, etc. According to BIEO tags, B-a is usually followed by I-a however cannot be followed by B-s or I-s. The input sentence is $X = (x_1, x_2, \ldots, x_n)$ and, $Y = (y_1, y_2, \ldots, y_n)$ represents the prediction results. The prediction result Y is:

$$\text{Score}(X, Y) = \sum_{i=1}^{n} P_{i,y_i} + \sum_{i=0}^{n} A_{y_i,y_{i+1}} \tag{6}$$

where P is the matrix of the scores output by the last layer, and A is a matrix of the transition scores; P_{i,y_i} is the score of the y_i^{th} tag of the i-th word in a sentence; $A_{y_i,y_{i+1}}$ represents the score of a transition from the tag y_i to y_{i+1}; n is the length of a sentence. The CRF model predicts labels for each word as mentioned above. All of the scores can be calculated according to the following formula:

$$P(y|X) = \frac{e^{S(X,y)}}{\sum_{y \in Y_X} S(X, y')} \tag{7}$$

In the final decoding, the optimal sequence labeling is computed as follows:

$$y^* = argmax\ S_{\widetilde{y} \in Y_X}(X, \widetilde{y}) \tag{8}$$

3.3. Educational Relation Extraction

As mentioned above, the main goal of this module is to identify the logical relationships that exist in educational entities and that facilitate learners learning more effectively. The above research [8–10,16,22] points out that these relationships are very important for learners: inclusion relationship, precursor relationship, identity relationship, sister relationship and correlation relationship. Table 2 shows the relationship categories and their definitions. The location of the entities is very important in determining the relationship. We learned from the R-BERT model to fuse the features of the sentence and educational entities, which are identified in the previous module.

Table 2. Educational relationships and their descriptions.

Relation Type	Relation Definition
Inclusion Relationship	Knowledge point A contains knowledge point B, and knowledge point B is the refinement of knowledge point A
Precursor Relationship	Knowledge point A must be learned before learning knowledge point B
Identity Relationship	Knowledge point A and knowledge point B are different descriptions of the same knowledge
Sister Relationship	Knowledge point A and knowledge point B have the same parent knowledge point C, and there is no learning sequence
Correlation Relationship	Knowledge point A and knowledge point B do not conform to the previous relationships, although they are still relevant

For a given sentence $S = \{c_1, c_2, \ldots, c_n\}$ with two target entities e_1 and e_2, we add a special token "#" at the boundary of the entity to locate the entity. We also insert "[CLS]" to the beginning of the input sentence. The output of the "[CLS]" can be used as a vector representation of the sentence. Suppose the final hidden state of BERT is M; for the final hidden state vector M_0 of the token "[CLS]", we added an activation operation and a fully connected layer. Its equation is:

$$M_0' = W_0[\tanh(M_0)] + b_0 \tag{9}$$

where $W_0 \in R^{d*d}$, and d is the hidden state size from BERT.

In addition to the sentence vector, we also need to combine the vectors of two entities. The entity vector is obtained by calculating the average value of each word vector. This process can be expressed as:

$$E = \frac{1}{j-i+1}\sum_{t=i}^{j} M_t \tag{10}$$

where i and j denote the beginning and end of the target entity.

For each entity vector, the process described in Formula (9) needs to be conducted. We concatenated two entity vectors and the vector of the token "[CLS]", then added a fully connected layer, which can be described as follows:

$$M'' = W_1\left[\text{concat}\left(M_0', M_1', M_2'\right)\right] + b_1 \tag{11}$$

where $W_1 \in R^{L*3d}$, L is the total number of relationship types. In this work, we used $L = 5$. M_1 and M_2 denote the final hidden state vectors of entity 1 and entity 2, respectively. In Equations (9) and (11), b_0, b_1 are bias vectors. Finally, we added a SoftMax layer for classification prediction.

3.4. Teacher Speech Fusion Module

Teacher speech is also a key resource in the educational knowledge graph, which is different from text and belongs to another modal resource. This module uses speech recognition technology to process audio signals and to integrate the teacher's speech into

the text knowledge graph. The main speech recognition process includes acoustic feature extraction, the conversion of the features into pronunciation of a phoneme sequence/pinyin sequence through the acoustic model, and the speech model transforms the phoneme sequence into text that a human can understand. The main process is shown in Figure 4.

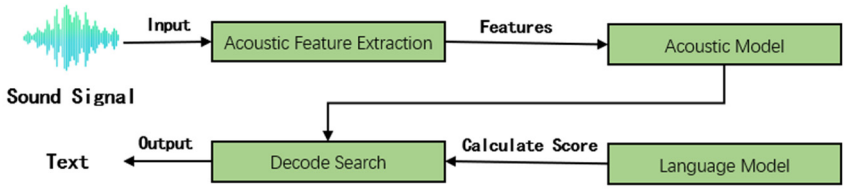


Figure 4. The main process of speech recognition.

We converted the audio data into WAV speech fragments that were able to be processed, obtained the frame number and sound channel of the sound, and used MFCC to extract the speech feature vector matrix to construct the input tensor. Then, the tensor needed to be framed and windowed, and the vector matrix was generated through the Fourier transform. The acoustic model used the deep learning framework Keras (https://keras.io/) (accessed on 8 February 2022) to define an 11-layer neural network CNN structure, and the loss function used Connectionist Temporal Classification (CTC) algorithm. CTC algorithm can realize end-to-end network training without the audio data being pre-aligned and requires only one input sequence and one output sequence. CTC algorithm can directly output the probability of the sequence prediction without external post-processing. Therefore, CTC algorithm is also used to decode the recognition results and to generate phoneme sequences. These phoneme sequences are used as the input of HMM language model, and the decoding process from phoneme sequences to text is realized through the language dictionary.

After completing speech recognition, we linked the speech entities with text-based education entities. To distinguish the relationship between these text entities, we redefined the relationship between speech and text entities as an "association relationship". First, the extracted set of educational entities was constructed as a domain lexicon. Each educational concept in the entity set has a unique id. Second, the speech entity is numbered. Then, the speech and the entities in the graph are linked by text matching technology to form a new triple grouping.

4. Results and Discussion

To evaluate our proposed construction system, we constructed an exemplary knowledge graph for the professional computer curriculum data structure. The performance of the system was evaluated comprehensively.

4.1. Experiment Settings

4.1.1. Dataset

To make our knowledge graph richer and more comprehensive, we created a dataset based on the data collected from curriculum teaching resources and online education resources. The online data mainly included course outline, Baidu entries, and Jianshu articles, the offline classroom data included courseware, textbooks, and teacher audio.

First, the domain experts marked the entities of the course outline (https://wenku.baidu.com/view/9ac023c85901020207409ce8.html) (accessed on 8 February 2022), resulting in a total of 233 entities. Based on the course outline, we trawled through unstructured educational text resources from the Baidu Encyclopedia and Jianshu books (each keyword gets the first 30 pages of articles). The beautifulSoup4 library (https://www.crummy.com/software/BeautifulSoup/) (accessed on 8 February 2022) and the Selenium library (https://www.selenium.dev/) (accessed on 8 February 2022) were used to crawl from the website. The Baidu Encyclopedia data set comprised a total of 15,674 sentences, and there were 6690 the Jianshu article in total. For the courseware, text data were extracted and

saved based on text type, obtaining 8793 sentences. To minimize the noise of the original data, data cleaning was required to remove irrelevant symbols from the text. In addition, we also downloaded and saved the entirety of the classroom audio. The collected audio resources were uniformly converted into WAV format for post-processing.

4.1.2. Data Preprocessing

To evaluate the entity recognition task, 8793 courseware sentences were obtained as labeled datasets. The educational concepts were labeled according to the BIES label set proposed above. The main objects are three types of educational entities: "/a" for algorithm, "/s" for structure, and "/c" for basic terminology. The labeled datasets are divided into training and test sets at a ratio of 7:3.

For the relation extraction task, the data that were extracted from the relationship came from courseware, Baidu entries, and Jianshu articles, and data such as graphs that were not relevant for our research were removed. In order to enable BERT to capture the location information of the two entities, a special tag [CLS] was added to the front of each sentence, and the sentences were separated by [SEP]. At the beginning and end of the two entities, the tag "#" was used to distinguish the entities in the sentence. When the model recognizes the first two "#", then it takes the data between the two tags as the first entity, and the data between the two tags as the second entity when it identifies the third and fourth "#". After the entities are marked, the relationship between the two entities in the sentence must also be marked after the sentence. For the courseware, if the sentence contains two entities, then it should be marked according to the above method. When the sentence contains more than one entity, then the sentence is discarded directly. In this paper, we manually labeled 1452 sentences in the courseware. For Baidu entries and Jianshu articles, only sentences containing two entities were retained, and sentences less than five words after word division were discarded. Jianshu articles were automatically labeled with the data labeled by three domain experts to expand the experimental training set. A total of 30,000 sentences were automatically labeled, and they were manually reviewed by domain experts. The sentences were divided into a training set and test set accordingly, at a 7:3 ratio.

4.1.3. Evaluation Metrics

Accuracy (*Acc*), Precision (*P*), Recall (*R*), and *F*1-Score (*F*1) were the evaluation criteria used for our experiments.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{12}$$

$$P = \frac{TP}{TP + FP} \times 100\% \tag{13}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{14}$$

$$F1\_score = \frac{2PR}{P + R} \times 100\% \tag{15}$$

where TP represents the number of labels that are positive and predicted to be positive. TN represents the number of labels that are negative and predicted to be negative. FP represents the number of labels that are negative and predicted to be positive. FN represents the number of labels that are positive and predicted to be negative.

4.2. Experiment Results on Entity Recognition

4.2.1. Model Comparison

For the proposed entity recognition model, the model parameters were first set during training. Batch size is the sample size used in each iteration, which can determine the

direction of the gradient descent. For the size of the dataset in this paper, the batch size was set to 64. The learning rate affects the convergence speed and fitting effect of the model. The learning rate was set to 0.0001 with the Adam optimizer, and the dropout rate was 0.5. We trained the model for 100 epochs using an early stopping strategy. For baseline models, we used the same parameters as those outlined in their original papers or during their original implementation.

To verify the recognition effect of the method in this paper, we compared our method against the results by baseline methods, including BiLSTM-CRF and BERT-BiLSTM-CRF. Table 3 shows all of the results of the baselines and our model.

Table 3. Comparison of experimental results of the BiLSTM-CRF model, BERT-BiLSTM-CRF model and our model. We ran all models 20 times and took the average. The results indicated statistical significance base on Student's *t*-test ($p < 0.05$).

Method	*P*	*R*	*F*1
BiLSTM-CRF	80.04%	82.07%	81.06%
BERT-BiLSTM-CRF	80.34%	85.49%	82.83%
EduBERT-BiLSTM-CRF	85.32%	85.72%	85.52%

The experimental results show that each model works well when using our dataset. BiLSTM combines a forward LSTM and a backward LSTM to model the information before and after the sentence in order to make better use of context information. The CRF layer can automatically learn the constraints in the sentence and can add constraint labels to the BiLSTM output, improving entity recognition performance. The precision and *F*1 values for BiLSTM-CRF are 80.04% and 81.06%, respectively. On the basis of BiLSTM-CRF, adding a BERT model can make better use of local and global information. According to the results, the model with BERT improved the *F*1 by 1.77%, indicating that BERT helps to improve the named entity recognition performance.

Due to the lack of databases in the education field, the traditional BERT model cannot extract the features in the sequence well. Educational data have their own characteristics that cannot be ignored. The results show that our model performs better and that both the precision and the *F*1 increased by 4.98% and 2.69%. One reason for this is that we used a fine-tuned BERT model. The fine-tuned BERT model provides domain awareness and enriches the context semantics of the vertical domain, which is particularly important for the vertical domain NER. Another main reason for this is that we used the lexicon method. By using a predefined domain dictionary to construct an adaptive embedding, more domain information is brought into the sequence labeling.

Based on the analysis of the experimental results, the method used in this paper has achieved good results. This paper also used this model to complete the entity prediction of unlabeled text data to ensure the quality of the entity data and of the knowledge graph.

4.2.2. Parameter Sensitivity 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 over-fitting 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 a model with a learning rate of 0.0001, the value of *F*1 was higher than that of the model 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 has better performance, so the dropout value of 0.5 was selected.

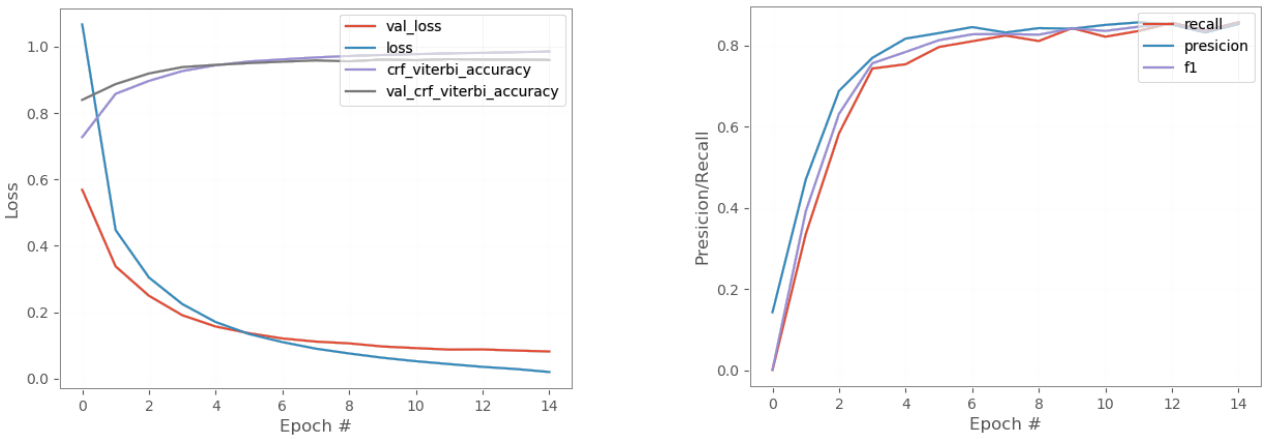
Table 4. Comparison of experimental results of our model with different learning rates. We ran all models 20 times and took the average. The results indicated statistical significance based on Student's t-test ($p < 0.05$).

Learning Rate	*F*1
0.01	82.43%
0.001	83.97%
0.0001	85.61%

Table 5. Comparison of experimental results of our model with different dropout values. We ran all models 20 times and took the average. The results indicated statistical significance based on Student's t-test ($p < 0.05$).

Dropout	*F*1
0.1	84.92%
0.3	85.31%
0.5	85.78%
0.7	85.12%
0.9	84.72%

Figure 5a shows that the loss value decreases gradually as the number of epochs increases, and it then decreases to a smaller and more stable value after 10 iterations. By comparing the above parameters, we selected the model with a learning rate of 0.0001 and a dropout value of 0.5. As shown in Figure 5b, this model had the best performance.



(**a**) (**b**)

Figure 5. (**a**) The loss value changes with the epoch; (**b**) model results with learning rate and dropout value of 0.0001 and 0.5, respectively.

4.3. Empirical Results on Relation Extraction

The results of all models are shown in Table 6. BiLSTM [42], CNN [29], and PCNN [41] were set according to their original papers. The experimental results show that the four models had some effect on the educational relation classification. It should be noted that the BiLSTM model performed the worst on our datasets. The reason for this could be that

although the BiLSTM uses two LSTMs to extract forward and reverse semantic features, as a result of the limited data, it cannot learn vocabulary features well. A vocabulary feature is an important factor in relation to extraction. CNN automatically extracts local features without complex data preprocessing. Using a single maximum pooling for convolution layer output can extract text feature representations to some extent, however it is difficult to capture the structural information between two entities. PCNN can divide the output of the convolutional layer into three parts based on the position information of the two entities, which could also be the reason why PCNN demonstrated better performance. Similarly, our model also incorporates entity location information. The difference is that our method uses the powerful coding capabilities of the BERT model and extracts both the semantic and grammatical features of the text. Our model achieved the highest accuracy with *F*1, achieving 75.26% and 80.39%, respectively.

Table 6. Comparison of experimental results of different methods. We ran all models 20 times and took the average. The results indicated statistical significance based on Student's *t*-test ($p < 0.05$).

Method	*Acc*	*P*	*R*	*F*1
BiLSTM	64.94%	67.49%	81.68%	73.91%
CNN	70.92%	76.25%	79.81%	77.99%
PCNN	73.80%	76.33%	81.45%	78.80%
Lexicon + R-BERT	75.26%	76.38%	84.85%	80.39%

In order to avoid over-fitting during model training, we considered the dropout parameters. Table 7 shows all the results of our model using different dropout values. Finally, we chose the dropout value of 0.5. The above experiments also prove the importance of entity information for the relation extraction. They also show that using special tags to mark entities can effectively input entity location information into the BERT model for training. BERT can obtain the grammatical and semantic features of sentences to classify relations more accurately. To ensure the quality of the knowledge graph built in this paper, the trained model is used to identify the relationship in the unlabeled data.

Table 7. Comparison of experimental results of Lexicon + R-BERT model with different dropout values. We ran all models 20 times and took the average. The results indicated statistical significance based on Student's *t*-test ($p < 0.05$).

Dropout	*F*1
0.1	79.14%
0.3	79.87%
0.5	80.39%
0.7	79.92%
0.9	78.67%

4.4. Visual Display of Knowledge Graph

The above experiments can be used to obtain the entities and relationships in the multi-modal educational knowledge graph. In the paper, the acquired educational concepts and relationships were stored in the Neo4j graph database (https://neo4j.com/) (accessed on 8 February 2022). Using D3.js (https://d3js.org/) (accessed on 8 February 2022) and the Vue framework (https://vuejs.org/) (accessed on 8 February 2022), a course multi-modal search system based on the knowledge graph was developed. The platform provides retrieval services based on multi-modal curriculum knowledge graphs, such as knowledge structure, learning sequence, and voice explanations. There are two main system functions: a course knowledge concept query module and a multi-modal concept display module.

4.4.1. Course Knowledge Concept Query Module

The search results obtained when using the "双向链表" (double-linked list) as an example are shown in Figure 6. The double-linked list was used as an entity to expand

and display the attribute information and the entity nodes related to it. With the help of the knowledge graph, the search no longer comprises of ordinary string matching, however, instead it consists of a semantic search that is based on the relationship in the graph. The returned knowledge graph can be generated dynamically based on the returned results, with nodes of different colors representing different entities. The arrows between the entities represent whether the relationship between entities is a one-way or two-way relationship. Learners can clearly understand the relationship between the knowledge points. The "链表" (linked list) and "双向链表" (double-linked list) in the figure represent a one-way predecessor relationship, which means that the double-linked list must be learned before the linked list knowledge points.

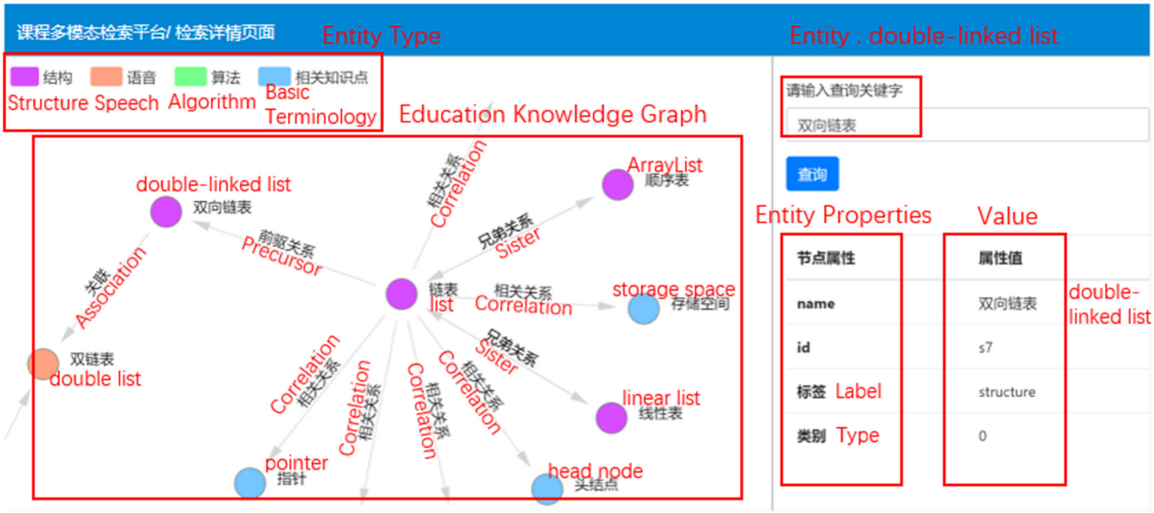


Figure 6. Course knowledge concept query module. Different colors represent different entity type. The arrows between entities represent the relationship between them. We use the "双向链表" (double-linked list) as an example. A portion of the education knowledge graph is shown on the left, and entity properties are displayed on the right.

4.4.2. Multi-Modal Concept Display Module

Figure 7 shows the results from when "广度优先遍历" (breadth-first search) was used as an example. When searching for this concept, the search results not only included text information, but also voice explanations. Learners can hear the teacher's voice in the classroom, which conforms to the learning mode that college students commonly engage in.

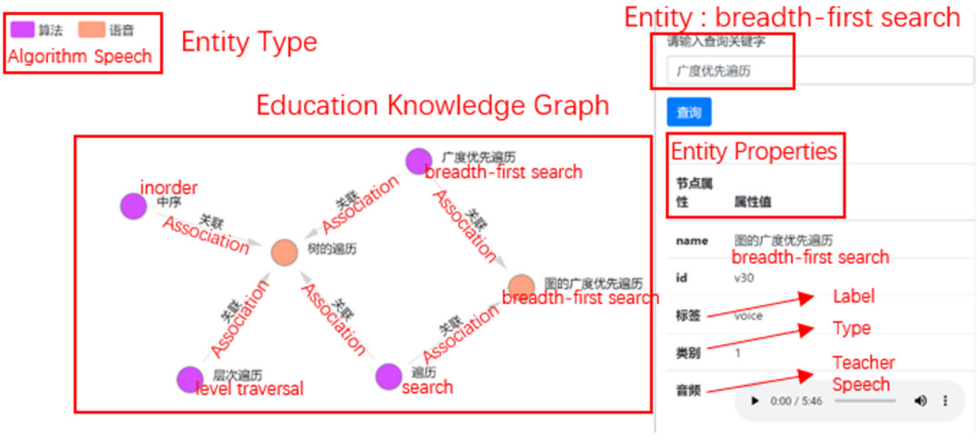


Figure 7. Multi-modal concept display module. When learners query breadth first traversal, they can not only see the text entities related to it, but they can also see their voice modal knowledge, that is, their teacher's voice.

5. Conclusions

In this work, we proposed a method for automatically constructing a multi-modal educational knowledge graph. It extracts the implied teaching concepts and educational relationships from heterogeneous data sources, most of which were online and offline real classroom teaching resources. While extracting the educational concepts, we also propose the introduction of domain knowledge into a fine-tuned BERT model by means of lexical enhancement. The educational relation extraction module combines the location information of the entities and BERT model to explore potential semantic relationships. We utilized speech recognition and text matching technology to embed teacher audio into the constructed text knowledge graph. The experimental results show that the entity recognition and relationship extraction precision have significant effects on our model. Finally, the multi-modal knowledge graph is constructed and stored in the neo4j database, which can be visualized by web programming.

In the future, we will explore whether such a multi-modal graph incorporating speech is more effective than one that does not include teacher speech. We will try to integrate more educational resources, such as course exercises and classroom videos, into the knowledge graph. Moreover, challenges related to multi-modal knowledge graphs, including named entity recognition task, will be looked at in more detail. Finally, we will focus on large-scale and high-quality multi-modal education knowledge graphs to provide better educational services for both teachers and learners.

Author Contributions: N.L.: Conceptualization, Methodology, Software, Data curation, Writing—Original draft preparation, Writing—Reviewing and Editing; Q.S.: Visualization, Writing—Reviewing and Editing; R.S.: Software, Validation; Y.C.: Investigation, Software; H.X.: Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62077027), the Ministry of Science and Technology of the People's Republic of China(2018YFC2002500), the Jilin Province Development and Reform Commission, China (2019C053-1), the Education Department of Jilin Province, China (JJKH20200993K), the Department of Science and Technology of Jilin Province, China (20200801002GH), and the European Union's Horizon 2020 FET Proactive project "WeNet-The Internet of us" (No. 823783).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data underlying this article are available in the article.

Acknowledgments: The authors would like to thank all of anonymous reviewers and editors for their helpful suggestions for the improvement of this paper.

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

References

1. Martín, A.C.; Alario-Hoyos, C.; Kloos, C.D. Smart Education: A Review and Future Research Directions. *Proceedings* **2019**, *31*, 57. [CrossRef]
2. D'Mello, S.K.; Olney, A.M.; Blanchard, N.; Samei, B.; Sun, X.; Ward, B.; Kelly, S. Multimodal Capture of Teacher-Student Interactions for Automated Dialogic Analysis in Live Classrooms. In *ICMI'15, Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*; Association for Computing Machinery: New York, NY, USA, 2015; pp. 557–566.
3. Suresh, A.; Sumner, T.; Jacobs, J.; Foland, B.; Ward, W. Automating Analysis and Feedback to Improve Mathematics Teachers' Classroom Discourse. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; 2019; Volume 33, pp. 9721–9728.
4. Anand, R.; Ottmar, E.; Crouch, J.L.; Whitehill, J. Toward Automated Classroom Observation: Predicting Positive and Negative Climate. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019; pp. 1–8.
5. Heiko, P. Towards Profiling Knowledge Graphs. In Proceedings of the PROFILES@ISWC, Vienna, Austria, 22 October 2017.
6. Wang, S.; Liang, C.; Wu, Z.; Williams, K.; Pursel, B.; Brautigam, B.; Saul, S.; Williams, H.; Bowen, K.; Giles, C.L. Concept Hierarchy Extraction from Textbooks. In Proceedings of the DocEng'15—ACM Symposium on Document Engineering 2015, Lusanne, Switzerland, 8–11 September 2015.

7. Lu, W.; Zhou, Y.; Yu, J.; Jia, C. Concept Extraction and Prerequisite Relation Learning from Educational Data. *Proc. Conf. AAAI Artif. Intell.* **2019**, *33*, 9678–9685. [CrossRef]
8. Hu, J.; Zheng, L.; Xu, B. An Approach of Ontology Based Knowledge Base Construction for Chinese K12 Education. In Proceedings of the 2016 First International Conference on Multimedia and Image Processing (ICMIP), Bandar Seri Begawan, Brunei, 1–3 June 2016; pp. 83–88.
9. Liu, H.; Ma, W.; Yang, Y.; Carbonell, J.G. Learning Concept Graphs from Online Educational Data. *J. Artif. Intell. Res.* **2016**, *55*, 1059–1090. [CrossRef]
10. Chen, P.; Yu, L.; 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 2018, London, UK, 26–28 June 2018.
11. Biega, J.A.; Kuzey, E.; Suchanek, F.M. Inside YAGO2s: A Transparent Information Extraction Architecture. In Proceedings of the 22nd International Conference on World Wide Web—WWW'13 Companion, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 325–328.
12. 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]
13. Fredo, E.; Günther, M.; Krötzsch, M.; Mendez, J.; Vrandečić, D. *Introducing Wikidata to the Linked Data Web*; Springer: Berlin/Heidelberg, Germany, 2014.
14. Liu, S.; Yang, H.; Li, J.; Kolmanič, S. Preliminary Study on the Knowledge Graph Construction of Chinese Ancient History and Culture. *Information* **2020**, *11*, 186. [CrossRef]
15. Su, Y.; Zhang, Y. Automatic Construction of Subject Knowledge Graph based on Educational Big Data. In Proceedings of the ICBDE'20—3rd International Conference on Big Data and Education, London, UK, 1–3 April 2020.
16. Yang, Y.; Liu, H.; Carbonell, J.G.; Ma, W. Concept Graph Learning from Educational Data. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China, 2–6 February 2015.
17. Senthilkumar, R.D. Concept Maps in Teaching Physics Concepts Applied to Engineering Education: An Explorative Study at The Middle East College, Sultanate of Oman. In Proceedings of the 2017 IEEE Global Engineering Education Conference, EDUCON 2017, Athens, Greece, 25–28 April 2017; pp. 107–110.
18. Chen, L.; Ye, L.; Wu, Z.; Pursel, B.; Lee, C.G. Recovering Concept Prerequisite Relations from University Course Dependencies. *Proc. AAAI Conf. Artif. Intell.* **2017**, *31*, 10550.
19. Kai, S.; Liu, Y.; Guo, Z.; Wang, C. Visualization for Knowledge Graph Based on Education Data. *Int. J. Softw. Inform.* **2017**, *10*, 3.
20. Zheng, Y.; Liu, R.; Hou, J. The Construction of High Educational Knowledge Graph Based on MOOC. In Proceedings of the 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC, Chengdu, China, 15–17 December 2017; pp. 260–263.
21. Dang, F.; Tang, J.-T.; Pang, K.; Wang, T.; Li, S.; Li, X. Constructing an Educational Knowledge Graph with Concepts Linked to Wikipedia. *J. Comput. Sci. Technol.* **2021**, *36*, 1200–1211. [CrossRef]
22. Yao, S.; Wang, R.; Sun, S.; Bu, D.; Liu, J. Joint Embedding Learning of Educational Knowledge Graphs. *arXiv* **2019**, arXiv:1911.08776.
23. Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; Dyer, C. Neural Architectures for Named Entity Recognition. NAACL. *arXiv* **2016**, arXiv:1603.01360.
24. Yadav, V.; Bethard, S. A Survey on Recent Advances in Named Entity Recognition from Deep Learning models. *arXiv* **2018**, arXiv:1910.11470.
25. Nadeau, D.; Sekine, S. A Survey of Named Entity Recognition and Classification. *Lingvist. Investig.* **2007**, *30*, 3–26. [CrossRef]
26. Fu, G.; Kit, C.; Webster, J.J. A Morpheme-based Lexical Chunking System for Chinese. *Int. Conf. Mach. Learn. Cybern.* **2008**, *5*, 2455–2460.
27. Wang, Y.; Wang, L.; Rastegar-Mojarad, M.; Moon, S.; Shen, F.; Afzal, N.; Liu, S.; Zeng, Y.; Mehrabi, S.; Sohn, S.; et al. Clinical information extraction applications: A literature review. *J. Biomed. Inform.* **2018**, *77*, 34–49. [CrossRef]
28. Sutton, C.; McCallum, A. An Introduction to Conditional Random Fields. *Found. Trends Mach. Learn.* **2012**, *4*, 267–373. [CrossRef]
29. Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; Kuksa, P.P. Natural Language Processing (Almost) from Scratch. *J. Mach. Learn. Res.* **2011**, *12*, 2493–2537.
30. Zhiheng, H.; Xu, W.; Yu, K. Bidirectional LSTM-CRF Models for Sequence Tagging. *arXiv* **2015**, arXiv:1508.01991.
31. Yann, D.; Fan, A.; Auli, M.; Grangier, D. Language Modeling with Gated Convolutional Networks. *Proc. Mach. Learn. Res.* **2017**, *70*, 933–941.
32. Hang, Y.; Deng, B.; Li, X.; Qiu, X. TENER: Adapting Transformer Encoder for Named Entity Recognition. *arXiv* **2019**, arXiv:1911.04474.
33. Hui, C.; Lin, Z.; Ding, G.; Lou, J.-G.; Zhang, Y.; Börje, F.K. GRN: Gated Relation Network to Enhance Convolutional Neural Network for Named Entity Recognition. *arXiv* **2019**, arXiv:1907.05611.
34. Zhu, Y.; Wang, G.; Börje, F.K. CAN-NER: Convolutional Attention Network for Chinese Named Entity Recognition. *arXiv* **2019**, arXiv:1904.02141.
35. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep Contextualized Word Representations. NAACL. *arXiv* **2018**, arXiv:1802.05365.
36. Radford, A.; Karthik, N. *Improving Language Understanding by Generative Pre-Training*; OpenAI: San Francisco, CA, USA, 2018.

37. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805.
38. Aydar, M.; Ozge, B.; Özbay, F. Neural Relation Extraction: A Survey. *arXiv* **2020**, arXiv:2007.04247.
39. Vu, N.T.; Adel, H.; Gupta, P.; Schütze, H. Combining Recurrent and Convolutional Neural Networks for Relation Classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; Association for Computational Linguistics: San Diego, CA, USA, 2016. [CrossRef]
40. He, H.; Ganjam, K.; Jain, N.; Lundin, J.; White, R.; Lin, J. An Insight Extraction System on BioMedical Literature with Deep Neural Networks. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics: Copenhagen, Denmark, 2017. [CrossRef]
41. Zeng, D.; Liu, K.; Chen, Y.; Zhao, J. *Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks*; EMNLP: Lisbon, Portugal, 2015.
42. Zhang, S.; Zheng, D.; Hu, X.; Yang, M. *Bidirectional Long Short-Term Memory Networks for Relation Classification*; Fujitsu Research and Development Center: Beijing, China, 2015.
43. Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The Graph Neural Network Model. *IEEE Trans. Neural Netw.* **2009**, *20*, 61–80. [CrossRef]
44. Zhang, Y.; Qi, P.; Manning, C.D. Graph Convolution over Pruned Dependency Trees Improves Relation Extraction. *arXiv* **2018**, arXiv:1809.10185.
45. Guo, Z.; Zhang, Y.; Lu, W. Attention Guided Graph Convolutional Networks for Relation Extraction. *arXiv* **2019**, arXiv:abs/1906.07510.
46. Fu, T.-J.; Li, P.-H.; Ma, W.-Y. *GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction*; Association for Computational Linguistics: Florence, Italy, 2019.
47. Wu, S.; He, Y. Enriching Pre-trained Language Model with Entity Information for Relation Classification. *arXiv* **2019**, arXiv:1905.08284.
48. Li, C.; Ye, T. Downstream Model Design of Pre-trained Language Model for Relation Extraction Task. *arXiv* **2020**, arXiv:2004.03786.
49. Mikolov, T.; Chen, K.; Corrado, G.S.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.