

Article

Not peer-reviewed version

Knowledge Point Entity and Relation Extraction based on Pre-training Model in Education Resources

Liu Cao , [Chengwei Huang](#) ^{*} , Qihao Wang , Xiaoming Zhu

Posted Date: 30 October 2023

doi: 10.20944/preprints202310.1914.v1

Keywords: knowledge graph construction; knowledge point extraction; pre-trained language model; relationship extraction



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Knowledge Point Entity and Relation Extraction Based on Pre-training Model in Education Resources

Liu Cao, Chengwei Huang *, Qihao Wang and Xiaoming Zhu

Zhejiang Laboratory, Departement of Intelligent Education, Keji Avenue, Yuhang Zone, Hangzhou 310030, China; ORCID ID: Chengwei Huang <https://orcid.org/0000-0001-9060-6361>

* Correspondence: huangcw@126.com

Abstract: Based on the teaching material of "Robotics" course, this paper studied the automatic knowledge graph construction, including knowledge points extraction and knowledge point relation extraction. We proposed a new method of extracting the first-level, second-level, and third-level knowledge points as well as their pre-requisite relations of knowledge points in the textbook. For the problem of insufficient knowledge of the pre-trained language model in the specific field, methods such as incremental pre-training and optimization of cost functions are employed to integrate subject knowledge into the pre-trained language model, thus improving its effectiveness. To overcome the problem that the traditional method of relationship extraction can not be applied directly to the extraction of teaching materials, a new scheme for knowledge point relationship extraction based on keyword relationship is proposed. The experimental data from textbooks shows that the F1 score of knowledge point extraction reaches 93%, considerably improved compared to the traditional model. Consequently, the knowledge point entity extraction and relationship extraction methods based on the pre-trained model can effectively extract structured information and facilitate the automatic construction of knowledge graphs.

Keywords: knowledge graph construction; knowledge point extraction; pre-trained language model; relationship extraction

1. Introduction

The United Nations Educational, Scientific Cultural Organization (UNESCO) released the report "Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development" in 2019. According to the study, it is crucial to change the field of educational ecology with the use of artificial intelligence technology. [23] However, in some countries, especially developing countries, there needs to be more discussion on educational artificial intelligence.

In [29], Zhong et al. believes that in solving the problem of educational bottlenecks, innovative education will bring an unprecedented large-scale revolution to the field of education. It is increasingly urgent to use artificial intelligence to empower daily teaching activities, such as personalized learning, intelligent teacher preparation, planning students' learning paths, and recommend learning resources. The intelligent application needs to be based on the knowledge graph of education resources. Therefore, how to construct the knowledge graph and extract knowledge points and related relationships from the total mass of teaching resource data has become a fundamental problem that needs to be solved urgently.

Much research is dedicated to completing knowledge extraction tasks, and traditional relation extraction methods are divided into pipeline methods and joint extraction methods. The joint extraction method is an end-to-end model implemented through a neural network. It realizes entity recognition and relation extraction at the same time and can well preserve the association between entities and relationships. However, the complexity of the model is high, and it takes work to achieve better performance. The pipeline method extracts relationships based on entity extraction results, which has better flexibility but ignores the contextual information between entities. Both methods have their advantages and disadvantages. Regarding education scenarios, we have proposed a new solution to the problems encountered.

In this study, a novel method for creating knowledge graphs of professional disciplines using pre-trained language models is proposed. In order to change the problem of map creation into the problem of knowledge point entity extraction and relationship extraction, we establish triplets in subject areas, which are formed of knowledge points in textbooks and precondition connections between knowledge points. Furthermore, we define the subject keywords, the indivisible entities related to the subject in the text part of the knowledge points in the education resources, representing the smallest knowledge unit of a subject. The contributions of this paper are as follows:

- We solved the specific problems encountered in the construction of knowledge graphs in the field of education and explored a complete solution covering knowledge point extraction and relation extraction. Moreover, we proposed an innovative hierarchical relationship extraction method based on knowledge graphs and pretrained models, which played an essential and fundamental role in constructing subject knowledge graphs.
- We proposed a data set in the field of education disciplines, including the education resources of more than 30 courses such as "Advanced Mathematics", "Robotics", "Robot Programming, and Application". After careful labeling by the teachers and students of the college, a high-quality original data set was obtained.

2. Related work

2.1. Entity Extraction

The recognition of named entities is one of the most fundamental tasks in natural language process. It aims to identify entities with specific meanings within unstructured text sequences and label them with names of persons, places, institutions, etc. There are many downstream tasks that utilize these entities with special meanings, including relation extraction, coreference resolution and dialogue question answering. Traditional methods of entity extraction include the use of LSTM [5], RNN [27], CNN [9], and their improved models for feature extraction, combined with the use of CRF for sequential modeling, training dependencies between labels, and then determining the optimal order of labels.

There have been a number of new NER schemes proposed in recent years. According to LIU there are three layers of entity information extraction: character representation, text encoding, label decoding, combined attention mechanism, and transfer learning [13]. In 2021, Li proposed a modularized interaction network model referred to as MIN (Modularized Interaction Network) [11]. The span information is combined with word-level dependencies at the same time, including the NER Module (NER Module), the Boundary Module (Boundary Module), the Type Module (Type Module), and the Interaction Mechanism (Interaction Mechanism). Through an interactive mechanism, joint training allows information to be shared between the two subtasks of boundary detection and type prediction. According to Asgari-Chenaghlu, images can be combined with text. In addition to providing stronger robustness, combining text features with image features can provide more reliable recognition of named entities [1].

Only a few studies have been conducted on the application of entity extraction methods to education. It has been proposed by Hu Huiting et al to recognize named entities in educational technology texts using a method that achieves high accuracy in a corpus of educational technology texts [6]. The method, however, failed to fully utilize the massive vertical field corpus and failed to offer a solution for labels that were truncated between text. Chen Xiaoling et al. developed a framework to implement a semantic knowledge extraction method for a wide range of entities and relationships between them [3]. Although this method optimized the loss function based on the specificity of the subject entities, its universality needed to be enhanced.

Transfer learning has led to the development of pre-trained language models such as ELMo [20], GPT [21], BERT [10], RoBERTa [14], and XLNet [26] which have achieved great success in NLP. Unsupervised learning is used in these methods. In many NLP tasks, they have achieved state-of-the-art results by learning semantic information from large amounts of text, creating vectorized representations of text, and generating vectorized representations of texts. Among these, BERT uses a bidirectional Transformer encoder to characterize textual features. By using this approach, it is possible to obtain dynamic vector representations at the token level, which improves the ability to represent vectors and is widely used. Thus, RoBERTa uses a greater number of data points, larger batches, and a

longer training period. By utilizing dynamic masking and removing next-sentence-prediction tasks, better performance can be achieved in the representation of text.

2.2. Relation Extraction

Relation extraction' refers to the extraction of semantic relationships between entities from a text source, which is an important component of information extraction, as well as a core technology and a major challenge in knowledge graph construction [17]. The goal of this project is to weave together a variety of entities and concepts identified by named entities to form a knowledge graph [7].

Currently, research methods focus primarily on the extraction of relations between entities appearing in a single sentence, aiming to extract relations between two entities in a single sentence [4,12,19]. As a result, many relationships will exist between multiple sentences or even in different chapters, and it is difficult to resolve the extraction problem using sentence-level relationship extraction alone. One of the objectives of the recent research work has been to achieve document-level relation extraction as well as to solve the problem of relation extraction across sentences. Using a structured attention mechanism, Nan proposed a latent structure optimization model (LSR) to generate task-specific dependency structures [18]. Non-local interactions between entities are captured by it. A labeled edge G CNN model has been proposed by Sahu, which captures local and nonlocal dependencies between sentences [24]. To solve the class imbalance problem of document relation extraction, Tan proposed an adaptive local loss [25]. In Zhang, semantic segmentation tasks were applied to the extraction of document relations, and the global dependencies between the triples were captured [28].

Often, knowledge point relationship extraction in the education field is characterized by low co-occurrence probabilities and cross text situations. The dependencies between different entities are difficult to capture. To solve the problem of document-level relationship extraction in the education field, we consider using a supervised text classification method. On the basis of domain expert knowledge data, we employ a supervised text classification method. As a result, it achieved better results when combined with the keyword entities derived by entity extraction to assist knowledge point entities in the classification learning process as a result of the powerful long-text semantic classification ability of Sentence Bert [22]. Finally, we are able to develop a better model for extracting relationshipsineducation.

3. Method

At the specific implementation level, there are unique problems in the field of education. It is not easy to achieve the optimal effect by simply using the pre-trained language model so that some points can be further optimized. It mainly includes three aspects. One is optimization oriented at the specific business level, such as integrating subject knowledge information into the pre-training model, performing loss scaling on labels that are difficult to identify in textbooks, and studying the optimization of label truncation problems. The second creatively proposes an extraction scheme that integrates keyword knowledge graphs in the relation extraction task of knowledge points. The third is deep learning training optimization, including data enhancement, FGM, and mixed precision training [15,16].

3.1. Entity extraction

Our entity extraction model adopts the main model structure of RoBERTa-Pointer-Network. First, each character in the input sentence is input into Roberta to obtain the feature vector representation of each character. Then the obtained word vector is input into BiLSTM, and the bidirectional expression of the text sequence containing context information is obtained through the input gate, the forget gate, and the output gate. Finally, the output of the BiLSTM layer is used as the input of the pointer network layer. Each character's label score in the sequence text is calculated and compared with the standard label, and the optimal label sequence is obtained using the dynamic optimization algorithm. We combine the practice in the field of education to make the following optimization plan.

3.1.1. Incremental Pre-training

Incremental pre-training makes the model more suitable for downstream tasks during the pre-training process by adding task-oriented data. The implementation ideas of incremental pre-training can be roughly divided into data preparation, data processing, model import, training, Etc. In order to improve the performance of the algorithm in downstream tasks, this paper adopts the method of the masked language model. It uses the corpus of robotics to carry out incremental pre-training so that the model can learn the knowledge of the subject domain in advance, which is more suitable for the data of this paper and downstream tasks.

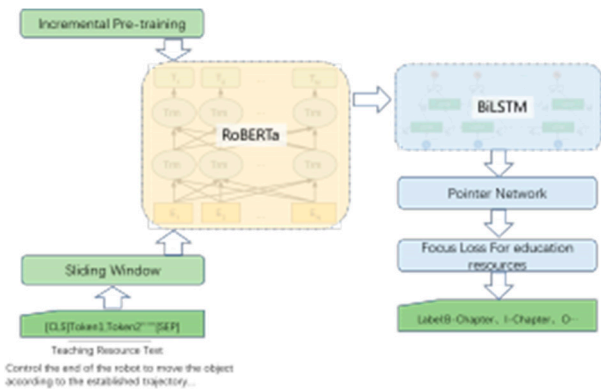


Figure 1. The main architecture of the entity extraction model.

Many experiments have proved that whole word masking is better than random masking because some masking in random masking is predicted based on lexical information, not based on the overall semantics, and we prefer the model to learn semantic information. Therefore, we built a professional vocabulary library in the field of education and used whole word masking for incremental pretraining.

Table 1. An example of the whole word masking.

Explanation	Sample
sentence	控制机器人末端以一定的夹持力 按既定轨迹移动物体
participle	控制 机器人 末端 以 一定的夹持力 按 既定 轨迹 移动 物体
random masking	[Mask]制机器人末端以一定的 [Mask]持力 按既定轨迹 [Mask]动物体
whole word masking	控制机器人末端以一定的 [Mask] [Mask] [Mask]按既定轨迹移动物体

3.1.2. Focus Loss For education resources

Due to the label extraction process of education resources, some labels are easier to extract, while others are more difficult to extract. For example, in teaching materials, the label of the chapter category is easier to extract, while the labels of sections and subsections are more challenging to identify. According to the difference in recognition difficulty, this paper designs a new type of loss function Focus Loss For education resources(FLFer). It makes the model pay more attention to difficult-to-classify samples during the training process, gives them higher weights, and reduces the impact of easy-to-classify samples on the total loss. The formula is as follows:

$$\alpha = [\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n], n \in \text{labels count} \quad (1)$$

$$\log(\text{pt}) = \begin{cases} \hat{p} & \text{if } y = \text{lable}, \\ 1 - \hat{p} & \text{otherwise} \end{cases} \quad (2)$$

$$\log(\text{pt}) = \log(\text{pt}) * \alpha \quad (3)$$

$$\text{Loss} = -(1 - \log(\text{pt}))^{\gamma} \log(\log(\text{pt})) \quad (4)$$

Among them, α is the loss of weight of each label. According to the difficulty of label learning, we set different weights for different labels and calculate the probability corresponding to each weighted label according to the Loss formula. This smoothly adjusts the weight ratio of easy-to-classify samples, balances the weights of difficult-to-classify samples, and reduces the weight of easy-to-classify samples. The model pays more attention to difficult-to-classify samples during training, and the more difficult to learn "specific" labels, the more significant the contribution to the loss value. This loss function can improve the practical effect of sections and subsections. See the ablation experiment section.

3.1.3. Sliding Window

When extracting knowledge points from education resources, there is a problem in that some labels cannot be extracted. Our analysis found that the data cut some entity nodes into two parts, and no data containing complete entities was passed into the model.

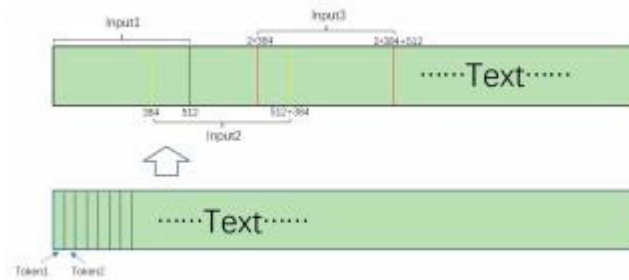


Figure 2. Model structure for sliding window.

We use the sliding window method in which the window length exceeds the step size, and model training is performed after data processing. The algorithm divides the original data set into a window of 512 characters with a step size of 384 to ensure that all entity characters can be completely present in the model input data. At the same time, it ensures a certain degree of semantic integrity and provides more comprehensive information for the model.

3.2. Relation Extraction

3.2.1. Prerequisite relation Extraction Method

We extract the prerequisite relations of the keywords based on obtaining the keywords. In the catalog of subject textbooks, the order of knowledge structure is defined as the title order of Chapter, Section, and Subsection. When keyword A appears in the title, if keyword A appears in the body of the subsequent title, and at the same time, the subsequent title also has the keyword B. According to the rules, the keyword A is considered the pre-keyword of the keyword B. According to this method, the front and rear maps of the keywords in the book can be obtained and expressed in the form of triples (head, relation, tail).

3.2.2. Model Architecture

As illustrated in Figure 3, the model consists of three parts: Knowledge point encoding layers (T-Encoder), prerequisite knowledge encoding layers (K-Encoder), and classification layers. The Knowledge point encoding model (T-Encoder) captures the lexical and syntactic information in the knowledge point. The prerequisite knowledge encoding model (K-Encoder) obtains the prerequisite relations of key- words in different knowledge points. The K-Encoder and T-encoder results are fused for classification in the final layer.

Given the dataset of knowledge point D , we randomly select two knowledge point $P = \{w_1, w_2, w_3, \dots, w_n\}$ and $Q = \{w_1, w_2, w_3, \dots, w_n\}$. The sequence of keyword $K_p = \{k_1, k_2, k_3, \dots, k_i\}$ and $K_q = \{k_1, k_2, k_3, \dots, k_j\}$ can be extracted from P and Q by the keyword extraction algorithm.

First, we encode the sentence-level contextual information for each knowledge point with T-encoder. For knowledge point $P = \{w_1, w_2, w_3, \dots, w_n\}$, the textual encode firstly sums the token embedding, segment embedding, positional embedding for each token to compute its input embedding, and the computes lexical and syntactic feature of knowledge point as follows:

$$SEP = \text{T-encoder}(w_1, w_2, w_3, \dots, w_n) \quad (5)$$

Do the same with Q :

$$SEQ = \text{T-encoder}(w_1, w_2, w_3, \dots, w_m) \quad (6)$$

Where T-encoder is a multi-layer bidirectional Transformer encoder. As T-encoder is identical to its implementation is RoBERTa and RoBERTa is prevalent, we exclude a comprehensive description of this module and refer reader to [14].

After encoding the knowledge point, our model adopts a knowledgeable encoder K-Encoder to inject the knowledge information into the representation of knowledge point. Given a sequence of key- words $K_p = \{k_1, k_2, k_3, \dots, k_i\}$, we represent each keyword k_i as

$$e_{k_i} = \text{K-encoder}(T) \quad (7)$$

Where the K-encoder is based on the effective knowledge embedding model GCN. We train the GCN model [8] based on the sets of prerequisite relation triplets (h, l, t) , where $h \in K_p$ and $t \in K_q$. A relationship $l \in L$ (the set of relationships). So the sequence of K_p and K_q can be represented as

$$EK_p = \{e_{k_1}, e_{k_2}, e_{k_3}, \dots, e_{k_i}\} \quad (8)$$

$$EK_q = \{e_{k_1}, e_{k_2}, e_{k_3}, \dots, e_{k_j}\} \quad (9)$$

To better integrate the different information for relationship, we aggregate the representation of knowledge point and the representation of keyword. For knowledge point P and Q , we can get the final the representation:

$$SK_p = SEP + EK_p \quad (10)$$

$$SK_q = SEQ + EK_q \quad (11)$$

Then, we use a FFNN with sigmoid function to calculate the probability of each relation:

$$P(r|P, Q) = \text{sigmoid}(W_r \begin{bmatrix} SK_p \\ SK_q \end{bmatrix} + b_r) \quad (12)$$

Where W_r , b_r are the weight matrix and bias for the linear transformation.

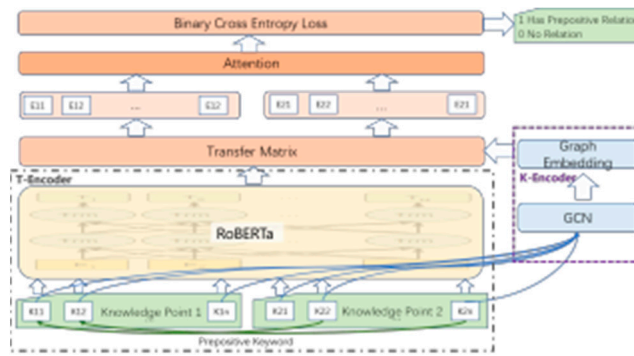


Figure 3. The main architecture of the relation extraction model.

3.3. Data Augmentation

Due to the high cost of manual annotation, the training dataset is limited. Under the labeled positive sample data, we use the method of data augmentation to construct complex negative samples, enhance the model ability, and reduce false recall. Each segment has labeled entity data after segmentation, and the negative sample is constructed by entity replacement. One is to replace the entity words with random words of equal length by randomly replacing tokens and constructing a negative sample without entity labels. The second is to use other entity words to replace and replace the correct entity words with other entity words in the label data to construct a negative sample that is difficult to recognize.

Each positive sample data corresponds to two negative samples, significantly enhancing the model's semantic discrimination ability and reducing false recall.

3.4. FGM

In order to enhance the dynamic representation ability of the pretrained model for text word embedding, this paper adds adversarial training in the process of word vector generation and model weight parameter training. The process of adding anti-disturbance in the Embedding layer is as follows:

Let the input sequence be $X = \{x_1, x_2, x_3, x_4, \dots, x_n\}$, where n is the sequence length of the input X . Input the input sequence into the RoBERTa model. The model performs vectorized calculation processing on the input X and inputs its word embedding calculation into the Transformer structure. This paper uses confrontation training when the pre-trained model is used for Embedding calculation. The formula is as follows:

$$r_{adv} = \epsilon \cdot g / \|g\|_2 \quad (13)$$

$$g = \nabla_x L(\theta, x + r_{adv}, y) \quad (14)$$

Where r_{adv} is the disturbance value, ϵ is the hyperparameter to control the disturbance size, g is the gradient of the solution, $\|g\|_2$ is the second norm of g , L is the loss function, Δ_x is the partial derivative for L , y is the label, θ is the model parameter. The generation of x'_i is centered on x_i and bounded by the L_p norm. The calculation steps are as follows:

1. For each x , calculate the forward loss and backpropagation of x to obtain the gradient
2. Calculate r_{adv} according to the gradient of the embedding matrix and add it to the current embedding, which is equivalent to $x + r_{adv}$ to generate an adversarial sample
3. Calculate the forward loss of $x + r$, backpropagate to get the gradient of the confrontation, and add it to the gradient of the first step
4. Restore embedding to the value in the first step
5. Update the parameters according to the gradient of the third step

The adversarial training can improve the model's defense ability against the adversarial samples and the generalization ability of the original samples. See the results of the ablation experiment.

4. Experimental Data

There are no publicly labeled textbook datasets in the education field. Since the field of education involves many subject areas, it is difficult to collect a total amount of data. We used the textbook data of more than 30 courses under the "Robotics" major, such as "Robotics," "Robot Operating Systems," "Robot Programming and Application," "Advanced Mathematics," Probability Theory and Mathematical Statistics, Etc. After careful labeling by the teachers and students of the college, a high-quality original data set was obtained. The data set includes more than 300 textbooks with 42,067,967 characters.

The original textbook data is in PDF format. We first convert the PDF format document into pictures page by page according to the page number. Then the picture is processed by cutting the header, and the header information on the top is removed to obtain a new picture. Finally, use the OCR engine to convert the image into text and generate text format data for data labeling and algorithm model training.

4.1. Label Definition

Text structure in the textbook can be reflected in the table of contents or the title of the main text. For example, most textbooks have first-level, second-level, and third-level titles, that is, chapters, sections, and subsections. Except for the table of contents and the beginning of textbooks, most textbooks will start with "Chapter One" as the beginning of the text, and then "Section One" will often appear according to the hierarchical structure. Below the first section level, there will be a "first subsection." For example, 1.1 and 1.1.1 usually represent the first section of the first chapter and the first subsection under the first section of the first chapter.

Keywords refer to the inseparable entities related to the subject in the textbook's text part of the subsection, representing the smallest knowledge unit of a subject. Since subsections are composed of text and contain a series of keywords, a set of keywords can characterize the characteristics of subsections.

4.2. Ontology Construction

Constructing the ontology layer of the subject map is mainly to sort out concepts in the field of education and define relationships, attributes, and related constraints. The core is the definition of schema, which requires a deep understanding of domain business. Therefore, this part is currently mainly sorted out by domain experts.

In the disciplinary knowledge graph in the field of education, the relationship between knowledge points includes the following three types: The superordinate relationship refers to the parent-child relationship, such as the containment relationship. In the textbook, chapters contain sections, sections contain subsections, and subsections contain several keywords. The prerequisite relation is the sequence of learning knowledge points, which can be used to plan and design the learning path in the learning process. The similarity relationship is the similarity between knowledge points, and the knowledge fusion of knowledge graphs can be carried out according to the similarity between knowledge points.

4.3. Data Annotation

After performing OCR processing on the original PDF teaching materials, we obtain a large amount of text. There was data cleaning performed, including line break removal, keyword deduplication, sequence number removal, stop words removal, full-width and half-width conversion, as well as English lowercase conversion.

Data labeling is performed by three labelers to ensure quality, and the following points should be taken into consideration: First, the labeling of entities should avoid the problem of nested entities. Efforts should be made to subdivide the labeling granularity as much as possible, and fine-grained entities should be labeled with priority. Second, some entities appear frequently, similar to stop words in general dictionaries. They may not be treated as entities in order to ensure that entities are

distinguished. Third, single Chinese characters may have different meanings when combined with different words, so they must be verified by experts before they are used as entities.

5. Experimental Results Analysis

5.1. Knowledge Point Extraction

The knowledge points in the education resources include the first-level, second-level, and third-level titles of the teaching materials. This article uses the experimental data set and the entity extraction model to extract entities from knowledge points.

The NER ablation experiments are as follows. Based on the original model, we incorporated incremental pre-training in the education field, optimized the loss function, and added a sliding window to make the model more capable of learning the education field’s knowledge characteristics and enhance the model’s generalization ability.

Table 2. Ontology Design of Educational Knowledge Graph.

Type	Category	Name	Abbreviation	Example
Entity	First-level knowledge point	Chapter	CHP	Chapter Four Sensors
Entity	Second-level knowledge point	Section	SET	4.1Classification of sensors
Entity	Third-level knowledge point	Subsection	SST	4.1.1Built-in sensors
Entity	Finest-grained knowledge point	KeyWord	KEW	Amount of displacement, Linearity
Relation	Hypernym relationship	Contain	COT	<Chapter Four Sensors,4.1Classification of sensors >
Relation	Prerequisite relationship	Preprositon	PPS	<4.1Classification of sensors,5.2Vision system >
Relation	Similar relationship	similarity	SML	<9.1Control technology,9.3 Feedback Control>

Table 3. The details of education resources dataset.

Label	Train	Val	Test
Chapter	35980	10280	5140
Section	33858	9674	4837
Subsection	22998	6571	3286
Keyword	313969	89541	44770
Prerequisite relation	171256	48930	24465

In order to reflect the advantages of adding modules to the method in this paper, six ablation experiments are set up:

1. Ablation focus loss for education resources;
2. Ablation incremental pre-training;
3. Ablation sliding window;
4. Ablation data augmentation;
5. Ablation FGM;
6. Ablation mixed precision training;

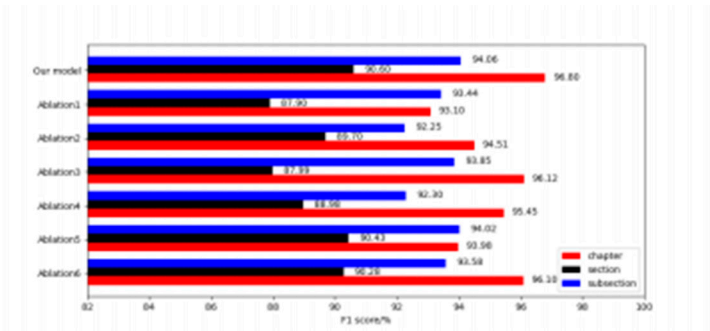


Figure 4. The ablation experiment of KP extraction.

It can be seen that the algorithm model proposed in this paper is more effective than the experimental ablation model on the knowledge point data set, with a maximum increase of 2.7%, 0.9%, and 0.81%, which proves the effectiveness of the algorithm module proposed in this paper. The loss function optimization has achieved good results. The reason for the analysis is that the characteristics of chapters are more pronounced than the other labels, which have general significance and are easy to identify. However, the section and subsection data are long and scattered, and different sections may have similar meanings, which are relatively difficult to identify. Therefore, good results have been achieved using the loss function proposed in this paper.

5.2. Relation Extraction

The goal of subject knowledge map construction is the front and back maps of knowledge points, and the front and back maps of knowledge points can be derived from the skeleton of keywords. Specifically, it is assumed that there are knowledge points KP1 and KP2, the relationship between KP1 and KP2, it can be obtained by incorporating the relationship between keywords.

Table 4. The details of the hyperparameters of the Textbook-Net.

Hyperparameters	Textbook-NER	Textbook-KP-RE
Training epoch	5	8
Batch size	32	32
Learning rate	5e-5	5e-5
Loss function	FLFer	Binary Cross Entropy Loss
α_{chapter}	0.8	/
α_{section}	1.1	/
$\alpha_{\text{subsection}}$	1.2	/
γ	0.9	/
Max length	512	512
Dropout	0.1	0.15
Mixed Precision	FP16&FP32	FP32

Keywords can represent knowledge points, and several keywords of the subject will appear in the text of specific knowledge points. We serialize the keywords and obtain the keyword graph. The keyword graph ontology structure is keyword A and keyword B and the prerequisite relation, which can be obtained through prior rules. In the catalog of subject textbooks, the order of the knowledge structure is defined, that is, the title order of chapters, sections, and subsections. When keyword A appears in a certain knowledge point, if the keyword appears in the text of a subsequent knowledge point, at the same time, keyword B also exists in the knowledge point, so keyword A is considered the pre-keyword of keyword B.

We use this priority rule to get more than 5,000,000 triple groups of keywords. Then we fuse the keyword graph with the extracted chapters, sections, and subsections and use the related model of knowledge graph embedding to train the embedding of keywords. For example, we used the TransE

[2]and GCN [8] models to train the relationship model between keywords and obtained the embedding after integrating the relationship.

We compares the effect experiments of different depth models and text fields as semantic labels. It verifies them on the expert annotation data set in robotics. We use the following knowledge graph embedding model for experimental comparison:

We order the optimal ROBERTa+GCN model as the Textbook-KP-RE model, which is uniformly used for semantic label comparison experiments. We use different corpora as semantic labels of knowledge points for experimental comparison. Among them, the name represents the name text of the knowledge point, the description represents the full description text of the subsection, and the keyword represents the keyword text extracted from the description text.

The results of semantic label experiments show that using the domain-optimized extraction model Textbook-KP-RE combined with knowledge point names and text keywords can achieve the best results, which is higher than other commonly used relationship extraction models.

Table 5. The experimental results (%) of KP relation extraction.

T-Encoder	K-Encoder	Precision	Recall	F1 score
BiLSTM	/	84.41	82.14	83.26
BiLSTM	TransE	86.43	85.15	85.78
BERT-base-chinese	/	89.53	88.89	89.2
BERT-base-chinese	TransE	89.23	89.01	89.12
BERT-base-chinese	GCN	91.32	91.1	91.2
ROBERTa-chinese	/	90.2	90.34	90.27
ROBERTa-chinese	TransE	92.58	91.89	92.23
ROBERTa-chinese	GCN	92.68	92.01	92.34

Table 6. The experimental results (%) of semantic label.

Semantic Label	Precision	Recall	F1 score
Description	84.17	82.22	83.18
KP-Name&Description	83.51	83.03	83.27
Keyword	89.44	91.19	90.31
KP-Name&Keyword	92.68	91.67	92.34

6. Conclusions

Aiming to solve the problems encountered in constructing the subject knowledge map, this paper, guided by the characteristics of the subject knowledge points and by the need to construct the subject knowledge map, investigated methods for knowledge extraction and relationship extraction represented by "Robotics," and proposed a Textbook-NER model as well as a Textbook-KP-RE model. According to the results of the experiments, the proposed method outperforms other models in terms of extraction accuracy and performance. In summary, the following conclusions can be drawn:

The first is the ontology layer construction of the subject knowledge graph, by which the conceptual architecture of the subject knowledge domain is defined. The second involves optimizing the loss function of the Textbook-NER model, as well as using extensive training and improvement programs. The third is that Textbook-KPRE model combines keyword extraction with relationship extraction and uses keyword granularity to extract prerequisite relations.

This paper intended to promote the construction of subject knowledge maps and contribute to subject development within the educational system. Thus, while we will continue to improve the subject knowledge graph, we will also continue to develop upper-level applications such as knowledge search, intelligent questions and answers, and intelligent recommendations based on the subject knowledge graph. By providing reference and service value to schools, teachers, and students, this will benefit the field of education in a number of ways.

References

1. Meysam Asgari-Chenaghlu, M Reza Feizi-Derakhshi, Leili Farzinvasht, MA Balafar, and Cina Motamed, 'Cwi: a multimodal deep learning approach for named entity recognition from social media using character, word and image features', *Neural Computing and Applications*, 1–18, (2022).
2. Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo, 'Multilingual knowledge graph embeddings for cross-lingual knowledge alignment', *arXiv preprint arXiv:1611.03954*, (2016).
3. Hu Y Cheng X, Tang L, 'Extracting entity and relation of landscape plant's knowledge based on albert model', *China Journal of Geo- information Science*, **23**(7), 1208–1220, (7 2021).
4. Tsu-JuiFu, Peng-Hsuan Li, and Wei-Yun Ma, 'Graphrel: Modeling text as relational graphs for joint entity and relation extraction', in *Proceedings of the 57th annual meeting of the association for computational linguistics*, pp. 1409–1418, (2019).
5. Sepp Hochreiter and Jürgen Schmidhuber, 'Long short-term memory', *Neural computation*, **9**(8), 1735–1780, (1997).
6. Lij HuH, 'Named entity recognition method in educational technology field based on bert', *China Computer Technology and Development*, **32**(10), 164–168, (10 2022).
7. Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip, 'A survey on knowledge graphs: Representation, acquisition, and applications', *IEEE transactions on neural networks and learning systems*, **33**(2), 494–514, (2021).
8. Thomas N Kipf and Max Welling, 'Semi-supervised classification with graph convolutional networks', *arXiv preprint arXiv:1609.02907*, (2016).
9. Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE*, **86**(11), 2278–2324, (1998).
10. JDMCKLee and KToutanova, 'Pre-training of deep bidirectional transformers for language understanding', *arXiv preprint arXiv:1810.04805*, (2018).
11. Fei Li, Zheng Wang, Siu Cheung Hui, Lejian Liao, Dandan Song, Jing Xu, Guoxiu He, and Meihuizi Jia, 'Modularized interaction network for named entity recognition', in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 200–209, (2021).
12. Kaihong Lin, Kehua Miao, Wenxing Hong, and Chaoyi Yuan, 'Relation extraction based on relation label constraints', in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pp. 2166–2170. IEEE, (2020).
13. Pan Liu, Yanming Guo, Fenglei Wang, and GuohuiLi, 'Chinese named entity recognition: The state of the art', *Neurocomputing*, **473**, 37–53, (2022).
14. Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov, 'Roberta: A robustly optimized bert pretraining approach', *arXiv preprint arXiv:1907.11692*, (2019).
15. Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, et al., 'Mixed precision training', *arXiv preprint arXiv:1710.03740*, (2017).
16. Takeru Miyato, Andrew M Dai, and Ian Goodfellow, 'Adversarial training methods for semi-supervised text classification', *arXiv preprint arXiv:1605.07725*, (2016).
17. Marie-Francine Moens, *Information extraction: algorithms and prospects in a retrieval context*, volume 1, Springer, 2006.
18. Guoshun Nan, Zhijiang Guo, Ivan Sekulić, and Wei Lu, 'Reasoning with latent structure refinement for document-level relation extraction', *arXiv preprint arXiv:2005.06312*, (2020).
19. S. Ning, F. Teng, and T. R. Li, 'Multi-channel self-attention mechanism for relation extraction in clinical records', *Chinese Journal of Computers*, **43**(5), 916–929, (2020).
20. Mark Neumann Peters, Matthew E and Mohit Iyyer, 'Deep contextualized word representations', in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 2227–2237, (2018).
21. Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al., 'Improving language understanding by generative pre-training', (2018).
22. Nils Reimers and Iryna Gurevych, 'Sentence-bert: Sentence embeddings using siamese bert-networks', *arXiv preprint arXiv:1908.10084*, (2019).
23. FENG Y C REN Y Q, WAN K, 'The sustainable development of artificial intelligence education—the interpretations and implications of artificial intelligence in education challenges and opportunities for sustainable development', *Modern Distance Education Research*, **31**(5), 3–10, (1 2019).
24. Sunil Kumar Sahu, Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou, 'Inter-sentence relation extraction with document-level graph convolutional neural network', *arXiv preprint arXiv:1906.04684*, (2019).
25. Qingyu Tan, Ruidan He, Lidong Bing, and Hwee Tou Ng, 'Document-level relation extraction with adaptive focal loss and knowledge distillation', *arXiv preprint arXiv:2203.10900*, (2022).

26. Z Yang, Z Dai, Y Yang, J Carbonell, RR Salakhutdinov, and XL-Net Le QV, 'generalized autoregressive pretraining for language understanding; 2019', Preprint at <https://arxiv.org/abs/1906.08237> Accessed June, 21, (2021).
27. Wojciech Zaremba, Ilya Sutskever, and Oriol , 'Recurrent neural network regularization', *arXiv preprint arXiv:1409.2329*, (2014).
28. Ningyu Zhang, Xiang Chen, Xin Xie, Shumin Deng, Chuanqi Tan, Mosha Chen, Fei Huang, Luo Si, and Huajun Chen, 'Document-level relation extraction as semantic segmentation', *arXiv preprint arXiv:2106.03618*, (2021).
29. SC Zhong, 'How artificial intelligence supports educational revolution', *China Educational Technology*, 3, 17–24, (2020).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.