2019 IEEE International Conference on Data Mining (ICDM)

Constructing Educational Concept Maps with Multiple Relationships from Multi-source Data

Xiaoqing Huang[1], Qi Liu[1,*], Chao Wang[1], Haoyu Han[1], Jianhui Ma[1], Enhong Chen[1], Yu Su[2], Shijin Wang[2]
[1]Anhui Province Key Laboratory of Big Data Analysis and Application
School of Computer Science and Technology, University of Science and Technology of China
{xqhuang, wdyx2012, hhy1212}@mail.ustc.edu.cn, {qiliuql, jianhui, cheneh}@ustc.edu.cn
[2]iFLYTEK Research {yusu, sjwang}@iflytek.com

***Abstract*—Concept map is an useful tool to help people organize and improve knowledge. Particularly in educational domain, it is beneficial for students and teachers to improve the learning and teaching quality. Traditionally, manual educational concept maps, provided by teachers, are quite time-consuming and limited to teachers' experience. Thus, it is meaningful to automatically construct high-quality concept maps. However, existing data-driven solutions only focus on either separate data source or single pedagogic relationship, which are not sufficient to satisfy actual demands. To this end, we propose a novel framework, named Extracting Multiple Relationships Concept Map (EMRCM), to construct multiple relations concept maps from Multi-source Data. Specifically, we design various targeted evidences to explore diverse information of multi-source data from different perspectives. Then, we employ three classic classifiers to bulid the predictive model for extracting key concepts and multiple concept relationships using the proposed evidences. We create a real dataset for empirically studying this problem. Extensive experiments on a real-world dataset show the effectiveness of our method.**

***Index Terms*—Educational concept map; Multi-source data; Multiple relationships;**

I. Introduction

Concept map, composed of various concepts and their relationships, is a widely-used graphical tools for organizing and representing knowledge [1]. Among diverse concept maps, educational concept maps concentrate on the pedagogic relationship between concepts. Thus, it is beneficial for students to organize and obtain knowledge of a subject. Figure 1 shows a real-world example of a concept map in the mathematics subject. Traditionally, concept maps were provided by teachers [2]. However, this approach is quite time-consuming and limited to teachers' experience. Fortunately, as the E-learning systems are becoming more and more widespread, which provide abundant question logs of examines and the content of textbooks, data-driven solutions can be applied to this task for automatically constructing high-quality concept maps.

There are some efforts in constructing relational concept maps in the literature. e.g., extracting Wikipedia concepts prerequisite relationships [3], [4]. Generally, the prior methods mainly focused on extracting the prerequisite relationship between concepts from single teaching materials. However, it still exists many unsolved problems due to the following considerations. First, to accurately extract concept maps, it is improper to concentrate on a single data source. Second, with thousands of key concepts describing different knowledges, the implicit relationship between concepts is much more complicated. In addition to construct learning sequence between concepts, it is also necessary to construct the concept closeness relationship [5]. To this end, we believe that it is necessary to extracting multiple relationships concept maps from multi-source data. Unfortunately, there are many technical and domain challenges along this line. First, multiple data sources contain diverse data types with different structures. So it is necessary to develop targeted treatments for each data source. Second, each data source contains multi-dimensional information, e.g., textbooks have both content information and structural information. Hence, it is essential to extract evidences from multiple types of views. Third, since the implicit relationship between concepts is convoluted, we have to design different methods for different kinds of relationships.

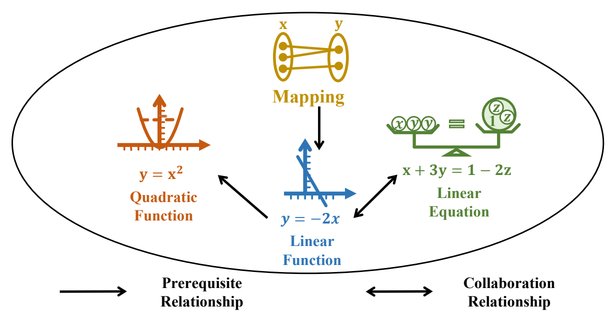

Fig. 1. The illustrative example of a concept map in the mathematics area.

To address the challenges mentioned above, in this paper, we develop a novel framework, named Extracting Multiple Relationships Concept Map (EMRCM), to construct multiple relations concept maps in large-scale online education systems by exploring different kinds of information from multi-source data. First, we aim at utilizing three different types of data to construct concept maps from different aspects, i.e., student question logs for private data, textbooks for authoritative data and Wikipedia for public data. Then, we attempt to figure out what kinds of information in multi-source data can be helpful to construct concept maps. Next, based on the meaningful features proposed above, we employ three different

2374-8486/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDM.2019.00132

IEEE computer society

binary classifiers to extract key concepts and build multiple relationships. Finally, we create a real dataset for empirically studying this problem[1]. Extensive experiments on a real-world dataset show the effectiveness of our method.

II. Related Work

In this section, we first introduce the general knowledge graph. Then, we focus on the educational concept maps construction.

Knowledge Graph. A knowledge graph organizes knowledge by linking entities with their relationships. It has been studied for a long time in the field of knowledge engineering. Prior work on the problem of identifying knowledge graphs inferred knowledge bases from a collection of noisy facts [6]. Recently, with the advent of Linked Open Data sources like DBpedia [7] , and by Google's announcement of the Google Knowledge Graph in 2012[2], representations of general world knowledge as graphs have attracted much attention again. There are various ways of building such knowledge graphs. They can be edited in a community-based way like Wikidata [8], or extracted from large-scale, semi-structured web knowledge bases such as Wikipedia [7] or YAGO [9].

In summary, relational facts are considered among existing word entities (e.g., "Trump") with a real-world relationship (e.g., "president") while our work pays more attention to virtual concepts (e.g., "equation") and the learning relationships between these concepts.

Concept Maps Construction. In our framework, one of the most important steps is to extract key concepts, and this is related to keyphrase extraction. This task aims at finding a small number of phrases to express the main topics of a document. TextRank [10] built an undirected and unweighted graph of the nouns or adjectives in a document and connected those that co-occur within a window of W words. Topic-based clustering methods such as KeyCluster [11] and TopicRank [12] aimed at extracting keyphrases that cover all the main topics of a document utilizing only nouns and adjectives and forming noun phrases that follow specific patterns. Talukdar et al. [3] and Liang et al. [4] studied prerequisite relationships by using Wikipedia articles. Based on some textbook features, Wang et al. [13] proposed a method to construct a concept map from textbooks, which jointly learns key concepts and their prerequisite relations.

However, these methods only consider either the teaching materials or student assessment data to extract prerequisite relationship, while our work extracts multiple relationships from multi-source data.

III. EMRCM Framework

In this section, we first give some preliminary definitions and then introduce technical details of EMRCM framework. Figure 2 demonstrates the workflow overview of our approach.

[1]https://github.com/xqhuang141/EMRCM-DATASET

[2]http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html

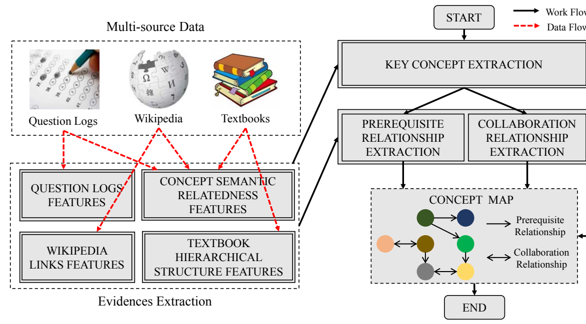

Fig. 2. An overview of the EMRCM framework.

A. Problem Definition

Definition 1 (Student question logs corpus). Student question logs $\mathcal{L}$ contains the scores and time costs of students on test questions, as well as the content, analysis and answers of each question. A question-answered record is denoted by a 5-elements tuple $(u, q, s_{uq}, t_{uq}, Con_q)$ which means that student $u \in U$ answers question $q \in Q$ with the score s_{uq} and the answer time t_{uq}. The score $s_{uq} = 1$ if the student u answers the question q correctly, otherwise, $s_{uq} = 0$. Con_q is the texts of question q which includes the question content $Con_{q^\dagger}$ and analysis $Con_{q^\ddagger}$.

Definition 2 (Subject textbooks corpus). Textbook corpus is composed by n textbooks in the same subject area, denoted as $\mathcal{S} = \{\mathcal{B}_1, \cdots, \mathcal{B}_i, \cdots, \mathcal{B}_n\}$, where $\mathcal{B}_i$ is one textbook. Each textbook $\mathcal{B}$ can be further represented as a subchapter set $\mathcal{B} = \{\mathcal{C}_1, \cdots, \mathcal{C}_i, \cdots, \mathcal{C}_{|B|}\}$, where $\mathcal{C}_i$ denotes the i-th subchapter in book $\mathcal{B}$. Finally, Each subchapter $\mathcal{C}$ contains titles and several sentences, denoted as $\mathcal{C} = \{ct, s_1, \cdots, s_i, \cdots, s_i\}$, where s_i is the i-th sentence of the subchapter.

Definition 3 (Wikipedia corpus). Wikipedia is the largest encyclopedia in the world and it contains massive pages denoted as $\mathcal{P} =< p_1, \cdots, p_i, \cdots, p_m >$, where p_1 is the i-th page. Each page $p = (p_t, p_{abs}, p_{con})$, where p_t, p_{abs}, p_{con} is the title, abstract and content in page p.

Problem Statement (Construct concept maps from multi-source data):

Given. Candidate concept set C, question logs corpus $\mathcal{L}$, subject textbooks corpus $\mathcal{S}$ and Wikipedia corpus $\mathcal{P}$;

Output. A concept map $\mathcal{G} = \{(w_1, w_2, r_1, r_2) | w_1, w_2 \in W, r_1, r_2 \in R\}$. W is the key concepts set. $R = \{0, 1\}$ mean that $r_1 = 0 \&\& r_1 = 0$ when w_1 and w_2 have no relationship; $r_1 = 1 \&\& r_2 = 0$ when w_1 and w_2 have prerequisite relationship; and $r_1 = 0 \&\& r_2 = 1$ when w_1 and w_2 have collaboration relationship; $r_1 = 1 \&\& r_2 = 1$ is not exist since w_1 and w_2 could have at most one type of relationship.

B. Extraction Evidences for Concept Maps Construction

In this section, we study how to mining multiple data to extract and combine evidences for concept maps construction.

1) Question Logs Features: Question logs corpus in exercises provides an important indication for constructing concept maps. First, a test question is designed to measure a students's

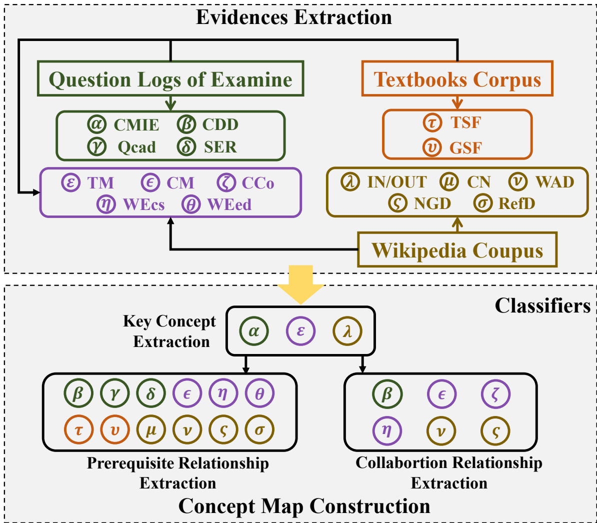

Fig. 3. An overview of evidences in concept map construction.

knowledge or skill[3]. Thus, there exist some key concepts in question content [14], [15]. Second, the difficulty of the question is closely related to the key concept in the question [16], [17]. Third, there are many concepts in the analysis of the questions, which is more likely to be the prerequisite of the concept in the questions content. Finally, student exercising records is also a good source of information [18], [19]. Therefore, in order to effectively utilizing the question logs corpus, we propose four novel features to help us construct the concept maps.

Content mention in questions. If content w is frequently mentioned by question content, then w is more likely to be a key concept. Based on this assumption, we propose this feature to extract key concepts. we define $CMIE(w) \in [0,1]$ as the normalized count mentioned in question content, as follows.

$$CMIE(w_i) = \frac{n_{w_i}}{\max\{n_{w_1}, \cdots, n_{w_i}, \cdots, n_{w_{|\mathcal{W}|}}\}},$$

where n_{w_i} is the number of the concept w_i appeared in question content.

Concept difficulty distance. Generally, the question difficulty refers to the percentage of the students answering the question correctly[4]. The difficulty of the question q is denoted as dif_q. The mean concept difficulty of the concept w_i is the average difficulty of all questions which contain the concept w_i, defined as follows.

$$CD(w_i) = \frac{\sum_{q\in\mathcal{L}} f(w_i, Con_{q^\dagger}) \cdot dif_q}{|\mathcal{L}|}.$$

where $f(w_i, Con_{q^\dagger})$ indicates the term frequency of the concept w_i in question content $Con_{q^\dagger}$, which reflects how important the concept w_i is in question q. The concept difficulty distance of the concept pair $\langle w_i, w_j \rangle$ is calculated as follow.

$$CDD(w_i, w_j) = CD(w_i) - CD(w_j).$$

[3]https://www.merriam-webster.com/dictionary/test
[4]https://www.assess.com/classical-item-difficulty-p-value/

Question content analysis distance. Given a concept pair $\langle w_i, w_j \rangle$, we propose the question content analysis weight to quantify the concept w_j mentioned in the question analysis record of concept w_i, defined as follows.

$$Qcaw(w_i, w_j) = \frac{\sum_{q\in\mathcal{L}} f(w_i, Con_{q^\dagger}) \cdot r(Con_{q^\ddagger}, w_j)}{\sum_{q\in\mathcal{L}} f(w_i, Con_{q^\dagger})},$$

where $r(Con_{q_2}, w_j)$ denoted whether concept w_j appears in question analysis Con_{q_2}. Naturally, if w_i appears in question content and w_j appears in question analysis, $Qcaw(w_i, w_j)$ tends to be larger, and $Qcaw(w_i, w_j) \in [0,1]$. The question content analysis distance($Qcad(w_i, w_j)$) is defined as follows.

$$Qcad(w_i, w_j) = Qcaw(w_j, w_i) - Qcaw(w_i, w_j).$$

Student exercising records. Before introducing this feature, let us define $Q(u)$ as the question set of student u and $\mathcal{I}(Q, w_i)$ as the question indexes that containS concept w_i in question set Q. For example, if w_i appears in the first and third question of Q, then $\mathcal{I}(Q, w_i) = \{1, 3\}$. In the answer sequence of student u, if the student makes a mistake in question with concept w_i, then the student is more likely to answer the wrong question with concept w_j. Based on this observation, for a given concept pair $\langle w_i, w_j \rangle$, we define $\mathcal{S}(Q) = \{(i,j) | i \in \mathcal{I}(Q, w_i), j \in \mathcal{I}(Q, w_j), i < j\}$ Suppose the concept w_i is the prerequisite of concept w_j. we calculate the feature of student exercising records as follows.

$$SER(w_i, w_j) = \frac{\sum_{u\in U} \sum_{(i,j)\in\mathcal{S}(Q)} s_{ui} - s_{uj}}{|U|},$$

where s_{ui} is the score of the student u in question i.

2) Textbook Hierarchical Structure Features: The table of contents (TOC) and the grade of the textbook indicate inherent relationships between concepts since teacher's curriculum planning is based on this information. In this section, we define two textbook hierarchical structure features, including TOC structure feature and grade levels of concepts feature, to help us infer relationships of concepts.

TOC structure feature. In real-world scenarios, there is no clear relationship between the chapters of the textbook. On the contrary, the subchapter in the chapter contains rich information for extracting concept relationships. For example, there is no clear relationship between the first chapter "algorithm" and the second chapter "statistics" in the second grade of high school textbook. However, the first subchapter "Probability of random events" and the second subchapter "classical models of probability" in chapter "probability" have a strong prerequisite relationship. Based on this line, we define the TOC structure feature as follows.

$$TSF(w_i, w_j) = \frac{\sum_{\mathcal{B}\in\mathcal{S}} \left(\sum_{\mathcal{C}\in\mathcal{B}} f(w_i, \mathcal{C}) - f(w_j, \mathcal{C})\right) / |B|}{|S|},$$

where $f(w_i, \mathcal{C})$ is the subchapter of chapter $\mathcal{C}$ which contains the concept w_i.

Grade structure feature. Grade structure feature is similar to the TOC structure feature but focuses on the grade level.

$$GSF(w_i, w_j) = \frac{\sum_{\mathcal{B}\in\mathcal{S}} f(w_i, \mathcal{B}) - f(w_j, \mathcal{B})}{|S|},$$

where $f(w_i, \mathcal{B})$ is the subchapter of textbook $\mathcal{B}$ which contains the concept w_i.

TABLE I
THE STATISTICS OF MATHEMATICS STUDENT LOGS DATASET.

statistics	Orignial	Pruned
# of records	1635976	1600416
# of exercises	1583991	1347538
# of students	19903	17381
Avg. exercises per student	82.2	92.1
Avg. concepts per question content	\	14.1
Avg. concepts per question analysis	\	23.2

3) Concept Semantic Relatedness Features: Given textbooks contents and Wikipedia corpus, we learn appropriate representations for concepts. To get a low-dimensional, continuous and dense semantic representation, we learn concept word embeddings [20] on textbooks contents and Wikipedia corpus. Let us denote v_{w_i} as the vector of the concept w_i. Based on this representation, we define the concept semantic relatedness features as follow.

Title match. Title is a summary of the content of the subchapter, indicating the main points in the subchapter. If a concept appears in the title, it is likely to be a key concept. Given a concept w_i and subchapter title ct,

$$TM(w_i, ct) = \begin{cases} 0 & \text{if } w_i \text{ is in ct} \\ 1 & \text{Otherwise} \end{cases}.$$

Concept match. Given a concept pair $\langle w_i, w_j \rangle$, if the concept w_i appears in the concept w_j, it is more likely that w_i has a prerequisite relationship with w_j. Generally, if $w_i \cap w_j \neq \varnothing$, it seems that they have a relationship.

$$CM(w_i, w_j) = \frac{\|w_i \cap w_j\|}{\max\{\|w_i\|, \|w_j\|\}}.$$

Concept co-occurrence. Count the co-occurrences of two concepts existing in a sentence from either a book subchapter or a Wikipedia page.

$$CCo(w_i, w_j) = \frac{\sum_{B \in \mathcal{S}} \sum_{\mathcal{C} \in \mathcal{B}} \sum_{s \in \mathcal{C}} r(s, w_i) \cdot r(s, w_j)}{\sum_{B \in \mathcal{S}} \sum_{\mathcal{C} \in \mathcal{B}} \sum_{s \in \mathcal{C}} r(s, w_i)},$$

where $r(s, w_i) \in \{0, 1\}$ is an indicator of whether concept w_i appears in sentence s.

Word embedding cosine similarity. Given a concept pair $\langle w_i, w_j \rangle$, semantic relatedness of two concepts can be reflected by their cosine similarity in the vector space.

$$WEcs(w_i, w_j) = \frac{v_{w_i} \cdot v_{w_j}}{\|v_{w_i}\| \cdot \|v_{w_j}\|}.$$

Word embedding euclidean distance. The euclidean distance of the concept w_i and the concept w_j in the vector space.

$$WEed(w_i, w_j) = \sqrt{\sum_{k=1}^{N} \left(v_{w_{ik}} - v_{w_{jk}}\right)^2}.$$

4) Wikipedia links features: Besides the information described above, Wikipedia which contains millions of page and page links are also very useful in detecting key concepts and concept relatedness [4].

In/Out degree for concept. This feature count In/Out Degree of the Wikipedia page for the concept w_i and the concept w_j. define as $\mathrm{In}(w_i)$, $\mathrm{Out}(w_i)$, $\mathrm{In}(w_j)$, $\mathrm{Out}(w_j)$.

Common neighbors of concept pair. The number of the common neighbors of concept pair $\langle w_i, w_j \rangle$.

$$CN(w_i, w_j) = \frac{\mathrm{Out}(w_i) \cap \mathrm{Out}(w_j) + \mathrm{In}(w_i) \cap \mathrm{In}(w_j)}{\max\{\mathrm{Out}(w_i), \mathrm{Out}(w_j)\} + \max\{\mathrm{In}(w_i), \mathrm{In}(w_j)\}}.$$

TABLE II
THE STATISTICS OF TEXTBOOKS DATASET.

statistics	Elementary	Middle	High
# of textbooks	12	6	8
# of subchapters	104	95	85
# of pages	1,400	1,093	1,237
# of key concepts	325	294	473
# of labeled pairs	1,576	1,374	962
# of prerequisite relations	258	208	352
# of collaboration relations	168	428	320

Wikipedia abstract definition. Concept w_i is likely to be w_j's prerequisite if w_i is used in w_j's Definition.

$$WAD(w_i, w_j) = \begin{cases} 1 & \text{if } w_i \text{ appears in } w_j\text{'s definition} \\ 0 & \text{Otherwise} \end{cases}.$$

Common normalized google distance. We compute the normalized google distance (NGD) of two concepts based on their Wikipedia links [21]. Define as follows.

$$NGD(w_i, w_j) = \frac{\max(\log|\mathrm{In}(w_i)|, \log|\mathrm{In}(w_j)|) - \log|\mathrm{In}(w_i) \cap \mathrm{In}(w_j)|}{\log N - \min(\log|\mathrm{In}(w_i)|, \log(w_j)|)}.$$

Reference distance. Liang et al. [4] proposed a new metric to measure the prerequisite relationships between concepts. If most related concepts of w_i refer to w_j, then w_j is more likely to be a prerequisite of w_i.

$$RefD(w_i, w_j) = \frac{\sum_{n=1}^{N} r(c_n, w_j) \cdot w(c_n, w_i)}{\sum_{n=1}^{N} w(c_n, w_i)} - \frac{\sum_{n=1}^{N} r(c_n, w_i) \cdot w(c_n, w_j)}{\sum_{n=1}^{N} w(c_n, w_j)}.$$

C. Concept Map Construction

There are many different kinds of classifiers that can be used in construct concept maps. Followed by Pan et al. [22], we employ three widely-used binary classifiers i.e.,Logistic Regression (LR), SVM with linear kernel (SVM) and Random Forest (RF) using different feature proposed in previous subsection to construct concept map. Specifically, as shown in Figure 3, we apply three novel classifier to extract concept map.

IV. EXPERIMENTS

In this section, we conduct extensive experiments to demonstrate the effectiveness of our method.

A. Experimental Dataset

In order to validate the efficiency of our features, we manually construct concept maps using three sections mathematics textbooks and student question logs which contains elementary, middle and high school. To the best of our knowledge, there is no public dataset for mining concept maps in textbooks which contain student question logs. we create the experimental data sets through a three-stage process.

The experimental dataset supplied by iFLYTEK Co., Ltd. is collected from Zhixue[5], a widely-used online learning system, which provides senior high school students with a large

[5]http://www.zhixue.com

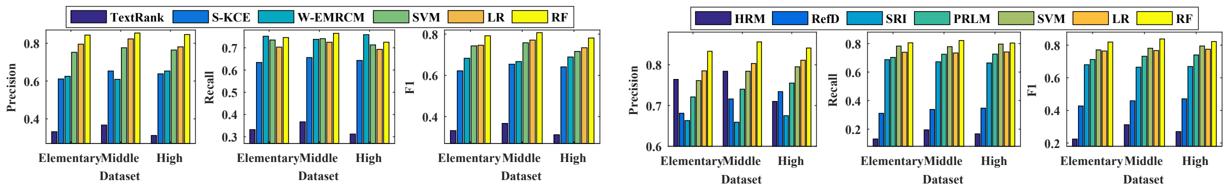

Fig. 4. Key concepts extraction and Prerequisite relationship extraction performance.

exercise resources for exercising To make sure the reliability of the experimental results, we filter the students that did less than 20 exercises and the exercises that no students have done. Table I shows the statistics of the dataset before and after preprocessing.

Then, for mathematics textbooks, we downloaded mathematics electronic textbooks for elementary, middle and high schools from the Internet, and then converted them to ".txt" format using the OCR tool. We manually labeled key concepts for each textbook, followed by Wang et al. [13], 1) Extract all Wikipedia concepts that appear in each book chapter. 2) Given a concept w_i, we select it as a candidate concept if $Titlematch(w_i, tw) = 1$ where tw is the title of subchapter, or w_i is ranked within top-20 among all concepts based on word embedding cosine similarity (WEcs) feature. 3) Label the candidates as "key concept" or "not key concept" and obtain a set of key concepts in this area.

Finally, we manually annotated the relationships among the labeled key concepts. For each concept pair $\langle w_i, w_j \rangle$, we manually labeled them as "w_i is w_j's prerequisite", "w_j is w_i's prerequisite", "w_i and w_j has collaboration relationship" or "no relationship". Table II shows characteristics of the datasets. For each dataset, three education experts with corresponding background knowledge were asked to label the data. We take a majority vote of the annotators to create final labels. We achieve an average 85% correlation for key concepts labeling task, an average 71% correlation for the key concept prerequisite relationships labeling task and an average 81% correlation for the key concept collaboration relationships labeling task.

As shown in Table II, there are much fewer positive instances than negative instances, so we balance the training set by oversampling the positive instances [23]. In our experiments, we employ 3 different binary classifiers, including Logistic Regression (LR), SVM with linear kernel (SVM) and Random Forest (RF). We use precision (P), recall (R), and F_1-score (F_1) to evaluate the EMRCM.

B. Key Concept Extraction Evaluation

1) Baseline Approaches: To investigate the model effectiveness, we compare the performance of our algorithm with several key concept extraction models, including TextRank [24], S-KCE [13] and W-EMRCM. Question log features are not used in W-EMRCM.

2) Performance Comparison: Figure 4 shows the performance of key concept extraction on three datasets. We find that Random Forest beats other classifiers, with best F1 across all three datasets. Moreover, our method with each classifier performs better than all other baselines. The results indicate that our feature can make full use of textbooks and question records. Compared with our method, W-EMRCM achieves high recall but low precision. This is because some key concepts may not appear in question content. But as long as it appears, it is more likely to be a key concept.

TABLE III
COLLABORATION RELATIONSHIP EXTRACTION PERFORMANCE

DataSet	Method	Pre	Rec	F_1
Elementary	SVM	0.732	0.756	0.744
	LR	0.786	0.725	0.754
	RF	**0.815**	**0.803**	**0.810**
Middle	SVM	0.694	0.712	0.703
	LR	0.729	0.703	0.716
	RF	**0.773**	**0.788**	**0.780**
High	SVM	0.728	0.746	0.737
	LR	0.766	0.714	0.739
	RF	**0.784**	**0.812**	**0.798**

C. Prerequisite Relationship Identification Evaluation

1) Baseline Approaches: To investigate the model effectiveness, we compare the performance of our algorithm with several prerequisite relationship extraction models, including HPM, RefD [4], SRI [13] and PRLM [22]. For HPM, we adopt the 10 lexico-syntactic Patterns used by [13] to identify prerequisite relationship.

2) Performance Comparison: Figure 4 shows the performance of prerequisite relationship identification on three datasets. There are several observations. First, we find that Random Forest also beats other classifiers, with best F1 across all three datasets. Second, our method with each classifier outperforms the baseline method across all three datasets. Moreover, we observe that our method achieves significantly higher precision (P), recall (R) and F_1 score than SRI. It demonstrates that question logs feature is very useful in the task of prerequisite relations extracting. Another potential reason is that our textbook hierarchical structure feature (THSF) contains both TOC information and grade structure feature. If two concepts are not in the same textbook (For example, textbook 1 and textbook 2), grade structure feature reveals that two concepts might have prerequisite relationships.

D. Collaboration Relationship Identification Evaluation

To the best of our knowledge, there is no previous work about collaboration relationship extracting, we only compare our method among different classifiers. As shown in Table III, the evaluation results varies from different classifiers. We find that Random Forest (RF) achieves best F1 across all

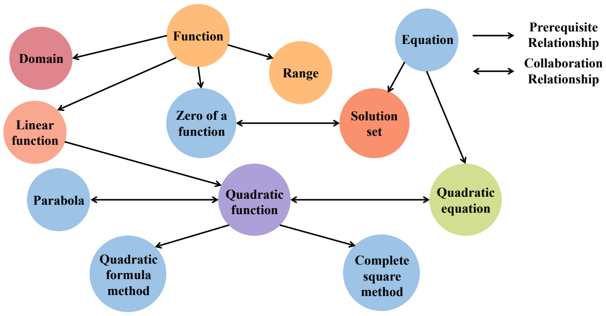

Fig. 5. The case study of concept maps contruction.

three datasets. The reason is as follows. Instead of a simple descriptive feature, each of our proposed feature determines whether a concept pair has collaboration relationship from a specific aspect; its function is similar to an independent weak classifier. Therefore, rather than using a linear combination of features for classification (e.g., SVM and LR), a boosting model (e.g., Random Forest) is more suitable for this task.

E. Case Study

In this section, we present a case study on concept map construction. From Figure 5, we can find that by considering both prerequisite and collaboration relationship, our method achieves better performance in concept map construction. For example, there is no clear learning sequence between "Quadratic function" and "Quadratic equation", but actually grasping one concept could better help us learn another concept. The concept map constructed by EMRCM contains both prerequisite and collaboration relationship, which are more advantageous and reasonable than others.

V. Conclusion

In this paper, we provided a focused study on multiple relationships concept map construction in the education domain. For modeling the multi-source of education data, we proposed several useful features from different aspects. Then, we extracted key concepts and multiple relationships from multi-source data to construct more reasonable and satisfactory concept maps. Finally, the experimental results on a large-scale real-world dataset clearly demonstrated the effectiveness of our model.

VI. Acknowledgement

This research was partially supported by grants from the National Key Research and Development Program of China (No. 2016YFB1000904), the National Natural Science Foundation of China (Grants No. 61922073, U1605251), and the Science Foundation of Ministry of Education of China & China Mobile (No. MCM20170507).

References

[1] J. D. Novak and A. J. Cañas, "The theory underlying concept maps and how to construct them," *Florida Institute for Human and Machine Cognition*, vol. 1, pp. 2006–2001, 2006.

[2] J. Stewart *et al.*, "Concept maps: A tool for use in biology teaching." *American biology teacher*, vol. 41, no. 3, pp. 171–75, 1979.

[3] P. P. Talukdar and W. W. Cohen, "Crowdsourced comprehension: predicting prerequisite structure in wikipedia," in *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*. Association for Computational Linguistics, 2012, pp. 307–315.

[4] C. Liang, Z. Wu, W. Huang, and C. L. Giles, "Measuring prerequisite relations among concepts," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 1668–1674.

[5] S.-Y. Teng, J. Li, L. P.-Y. Ting, K.-T. Chuang, and H. Liu, "Interactive unknowns recommendation in e-learning systems," in *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018, pp. 497–506.

[6] D. B. Lenat, "Cyc: A large-scale investment in knowledge infrastructure," *Communications of the ACM*, vol. 38, no. 11, pp. 33–38, 1995.

[7] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "Dbpedia: A nucleus for a web of open data," pp. 722–735, 2007.

[8] D. Vrandečić and M. Krötzsch, "Wikidata: a free collaborative knowledge base," 2014.

[9] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: a core of semantic knowledge," in *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007, pp. 697–706.

[10] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," in *Proceedings of the 2004 conference on empirical methods in natural language processing*, 2004.

[11] Z. Liu, P. Li, Y. Zheng, and M. Sun, "Clustering to find exemplar terms for keyphrase extraction," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics, 2009, pp. 257–266.

[12] A. Bougouin, F. Boudin, and B. Daille, "Topicrank: Graph-based topic ranking for keyphrase extraction," in *International Joint Conference on Natural Language Processing (IJCNLP)*, 2013, pp. 543–551.

[13] S. Wang, A. Ororbia, Z. Wu, K. Williams, C. Liang, B. Pursel, and C. L. Giles, "Using prerequisites to extract concept maps fromtextbooks," in *Proceedings of the 25th acm international on conference on information and knowledge management*. ACM, 2016, pp. 317–326.

[14] R. Wu, Q. Liu, Y. Liu, E. Chen, Y. Su, Z. Chen, and G. Hu, "Cognitive modelling for predicting examinee performance," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[15] Y. Yin, Q. Liu, Z. Huang, E. Chen, W. Tong, S. Wang, and Y. Su, "Quesnet: A unified representation for heterogeneous test questions," *arXiv preprint arXiv:1905.10949*, 2019.

[16] Z. Huang, Q. Liu, E. Chen, H. Zhao, M. Gao, S. Wei, Y. Su, and G. Hu, "Question difficulty prediction for reading problems in standard tests," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[17] Q. Liu, Z. Huang, Z. Huang, C. Liu, E. Chen, Y. Su, and G. Hu, "Finding similar exercises in online education systems," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 1821–1830.

[18] Y. Chen, J. P. González-Brenes, and J. Tian, "Joint discovery of skill prerequisite graphs and student models." *International Educational Data Mining Society*, 2016.

[19] Y. Su, Q. Liu, Q. Liu, Z. Huang, Y. Yin, E. Chen, C. Ding, S. Wei, and G. Hu, "Exercise-enhanced sequential modeling for student performance prediction," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[21] I. H. Witten and D. N. Milne, "An effective, low-cost measure of semantic relatedness obtained from wikipedia links," 2008.

[22] L. Pan, C. Li, J. Li, and J. Tang, "Prerequisite relation learning for concepts in moocs," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017, pp. 1447–1456.

[23] M. A. Yosef, J. Hoffart, I. Bordino, M. Spaniol, and G. Weikum, "Aida: An online tool for accurate disambiguation of named entities in text and tables," *Proceedings of the VLDB Endowment*, vol. 4, no. 12, pp. 1450–1453, 2011.

[24] O. Medelyan, I. H. Witten, and D. Milne, "Topic indexing with wikipedia," in *Proceedings of the AAAI WikiAI workshop*, vol. 1, 2008, pp. 19–24.