

Prerequisite Concept Maps Extraction for Automatic Assessment

Shuting Wang
Computer Science and Engineering
Pennsylvania State University
sxw327@cse.psu.edu

Lei Liu
HP Labs, Palo Alto, CA, USA
lei.liu2@hp.com

ABSTRACT

Traditional assessment modes usually give identical set of questions to each student, thus are inefficient for students to fix their problems. In order to perform an efficient assessment, we utilize prerequisite concept maps to find students' learning gaps and work on closing these gaps and proposed a two-phase model for concept map construction. Experiments on concept pairs with prerequisite relationships which are manually created show the promise of our proposed method. In order to meet the challenge of using concept maps in automatic assessments, we also derive a *top-k* concept selection algorithm which allows students to view different numbers of concepts.

1. INTRODUCTION

The increasing growth of massive online education resources promise new possibilities for educational tools and services. Assessment, which is the task of assessing learning achievements and providing feedbacks to learners is crucial to academic success and has been rapidly changed recent years [2]. Concept map, which provides a clear view of knowledge structure, is widely used in assessments [1, 6]. In this paper, we develop an automatic personalized assessment learning system to improve the efficiency for students' problems fixing using concept maps. We utilize a prerequisite concept map, which represents domain concepts and their learning dependencies [4], to detect and fix each student's own knowledge gaps. For instance, a student does not understand "Multiplication" might because he fails to know how to do "Addition", which is a prerequisite for "Multiplication". In this situation, instead of keeping the student working with "Multiplication", a better idea may be move him backwards by re-learning "Addition".

A running example of using prerequisite concept maps for assessment is shown in Figure 1. Students can start with any concept in the concept map, for instance, "Edge". *LearningAid* will provide student questions about rectangle. If the student gives the correct answer, which means

that he probably understand this concept, the system will recommend him to move forward along the prerequisite relationship, i.e., moving to the subsequent concepts of "Edge", for example, "Rectangle". When the student accepts the recommendation, he will answer questions about "Rectangle". If the "Rectangle" question is correctly answered, the system will repeat the process above, i.e., recommending the subsequent questions of "Rectangle". Otherwise, if the question is not correctly answered, it means that the student does not understand the concept "Rectangle" well and probably he also has problems with prerequisites of "Rectangle". Therefore, the system will recommend some prerequisite of "Rectangle" for the student to study, for example, "Right Angle". The system then repeats process describe above to facilitate assessments. By automatically recommending next step for students based on the current assessment results, our assessment systems with prerequisite concept maps offer personalized assessment learning experience. In summary, the system makes the following novel contributions:

- We propose a novel automatic personalized assessment system which finds students' learning gaps and work on closing the gaps using prerequisite concept maps.
- We investigate a two-phase model for prerequisite concept map construction which includes domain concepts extraction and prerequisite relationships identification.
- We deliver the concept map to students in a hierarchical way and allow them to view concepts in different levels.

2. PREREQUISITE CONCEPT MAP CONSTRUCTION

To automatically construct prerequisite concept map, we propose a two-phase method which extracts domain key concepts from expert created educational resources such as textbooks and papers, and then identifies prerequisite relationships between extracted concepts.

2.1 Key Concept Extraction

Given a domain, our goal is to exact key concepts related to this domain from educational resources. In this paper, we use Wikipedia to help the key concept extraction and enrich article content. We first construct a domain specific concept dictionary in which each concept is the title of a domain related Wikipedia page. Then given an article, we identify all Wikipedia concepts in the article using this dictionary and obtain a list of Wikipedia candidates. At last we select *top-k* concepts using the following features:

titleMatch: A Wikipedia concept is likely to be a key concept in an article if its title appears in the article's title.

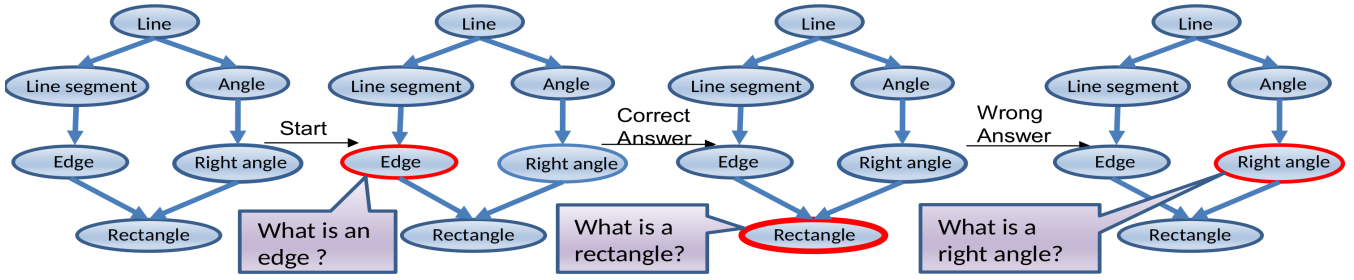


Figure 1: A running example for using prerequisite concept map to help student fix learning gap.

cosineSim: A Wikipedia concept is likely to be a key concept in an article if it has similar lexical contents with the article. $cosineSim$ captures the cosine similarity between the concept vector of Wikipedia candidate and that of the article.

Therefore, the domain key concept set consists of the top- k candidates based on $cosineSim$ score and those candidates with $titleMatch(\cdot)$ score equals to 1.

2.2 Prerequisite Relation Identification

Previous works mainly focus on hyponym relationship inference, which does not serve education purpose appropriately. Here we argue that there is likely to be prerequisite relationship between concept A and B if:

Usage of one concept in the definition of the other concepts: If concept A is used in B 's definition, A is likely to be B 's prerequisite.

Similar Content and Different Learning Levels: If two concepts cover similar topics, it is likely that they have some learning dependencies. For instance, "Network congestion" and "TCP congestion-avoidance algorithm" share a lot of topics such as packet loss and additive increase/multiplicative decrease, and "TCP congestion-avoidance algorithm" depends on "Network congestion". However, "TCP congestion-avoidance algorithm" and "Network security" do not share a lot of common topics and it is unlikely that there is a prerequisite relationship between them.

However, not all pairs of concept with similar content have prerequisite relationships. For example, "Transmission Control Protocol" and "User Datagram Protocol" cover similar topics while they are concepts at equivalent level of learning. Therefore, given two concepts, it is necessary to identify whether they are at different learning levels for prerequisite relationship inference.

Base on above arguments, we investigate three sets of features given pair of concepts A and B :

Usage in Definition Feature Feature $usage(\cdot)$ captures whether a concept is used in another concept's definition. $usage(A, B) = 1$ if A appears in B 's definition. A challenge is to obtain definitions of Wikipedia concepts. Here given a Wikipedia concept, we use the first sentence in its Wikipedia page as its definition. The reason of doing this is that most Wikipedia pages have a unified pattern with their first sentences as "*concept is definition of the concept*". For instance, Wikipedia concepts "Logarithm" and "e (mathematical constant)" are used to define "Natural logarithm" (definition for "Natural logarithm") is "The natural logarithm of a number is its logarithm ... " and $usage(logarithm, naturallogarithm) = 1$. In this case, "Logarithm" is a prerequisite of "Natural Logarithm".

Content Similarity Feature This feature captures the lexical similarity between Wikipedia concepts. We use feature $cosineSim(\cdot)$ defined in Section 2.1 to capture the content similarity between A and B .

Learning Level Features Learning level features measure whether a concept has a lower learning level and should be learned first. In order to capture this more precisely, we investigate three features to calculate the learning level of a concept: range of topic coverage, number of in-links and number of out-links. Given a collection of Wikipedia concepts C and a Wikipedia concept c_i , we define the following measurements:

Range of topic coverage: We first measure the learning level of a concept based on the range of topics covered by the concept [3]. Essentially, the more topics that a concept covers, the more basic the concept is. For instance, "Computer Network" covers more topics than "TCP" does and has a lower learning level and should be learned first. The range of topic coverage score $tc(c_i)$ of c_i is the Shannon Entropy over topics discussed in c_i . To be more specific, we run a topic model on C to generate s topics and the topic distributions for each concept in C . A topic model generates a document-topic matrix $F_{n \times s}$ where n is the number of concepts in the corpus. F_{im} , where F_{im} , with $i < |C|$ and $m < s$, is the probability that topic m is assigned to c_i . Then we compute the Shannon Entropy $H(c_i)$, i.e., the range of topic coverage score $tc(c_i)$, based on $F_{n \times s}$ as follows:

$$tc(c_i) = H(c_i) = \sum_m -F_{im} \log(F_{im}). \quad (1)$$

Number of in-links/out-links received: Besides the content information, millions of cross-page links in Wikipedia is also useful in detecting concept learning levels. If c_i receives a lot of in-links from other concepts, it is likely that c_i is fundamental in C and should be learned first. Similar conclusion can be drawn on the number of out-links of a concept.

Based on features defined above, for each concept, we then normalize value of each measurement to $[0,1]$ range, and take their weighted sum as the learning level score of the concept. The higher the learning level score is, the more basic the concept is.

We then investigate a threshold-based algorithm which selects pairs of concepts with prerequisite relationships. Given concepts A and B , if $usage(A, B) = 1$, we consider A to be the prerequisite of B and the algorithm stops; Otherwise, if the learning level difference and the content similarity between A and B are greater than some thresholds, A is considered as a prerequisite of B .

3. PREREQUISITE CONCEPT MAP DILIVERY

One challenge in assessment system using prerequisite concept map is that there might be hundreds of concepts in a domain. In this case, how can the system deliver the concept map to students? To meet these challenges, we investigate a *top-k* concept selection algorithm where k is a user specified parameter. In this case, the system constructs a subgraph of the original concept map which consists of k concepts in the domain where students can zoom in or zoom out to view different numbers of concepts. The rule of selecting *top-k* concepts is to choose domain important concepts while preserve connectivity of the subgraph. Given a Wikipedia concept A , we define A 's importance I_A as the similarity score between A 's contents and article's content. For the connectivity of the extracted subgraph $G = (V, E)$, where V is the concept set and E is the edge set in G , we use its edge density D as a measurement for its connectivity, which is defined as $\frac{2*|E|}{|V|*(|V|-1)}$.

The algorithm selects *top-k* concepts by maximizing the weighted sum of overall concept importance and subgraph edge density which is defined as $\arg \max_V (\alpha \sum_{k=1}^{|V|} I_k + (1 - \alpha)D(V))$, where α is the weight of concept importance and $1 - \alpha$ is the weight of subgraph connectivity.

The procedure of the algorithm is described as below:
 1) Initialize the concept set $V^{(0)}$ using *top-k* important concepts in the article. 2) For each iteration, concept set $V^{(n+1)} = V^{(n)} \setminus \{s\} \cup \{c\}$, where the replacement of concept s by concept c maximizes the value of objective function. 3) Iteratively update the concept set until there is no replacement that could increase the value of objective function.

4. EXPERIMENTS

4.1 Experiments Setup

In order to evaluate the extracted pairs of prerequisite relationship, we manually create dataset using a mathematics textbook¹ and a big data textbooks². We first extract domain key concepts from the book following the methods proposed in [5] by extracting candidate concepts from each book chapter using feature *titleMatch* and *cosineSim* and manually label each concept as "Important" and "Unimportant". Then we randomly select N concepts from the key concepts and manually label its subsequent and prerequisite concepts. Then we use this dataset as ground truth and report the F-1 score of our extracted concept map.

To affirm the effectiveness of our proposed learning level features, we propose a baseline which only uses features *usage in definition* and *Content Similarity* to identify concept relationships between concepts.

4.2 Experiment Results

Tables 1 and Table 2 show the accuracy of extracted concept relationships using different thresholds of concept relatedness and learning level. The column title is the threshold of concept relatedness and the row title is that of concept learning levels. Take "60% 80%" in Table 1 as example, this

¹Stewart, James, Lothar Redlin, and Saleem Watson. Pre-calculus: Mathematics for calculus. Cengage Learning, 2015.

²McKinsey Global Institute .Big data: The next frontier for innovation, competition, and productivity

means that if we select pairs of concepts with relatedness score higher than 60 percent of relatedness scores between all pairs of concepts and learning level difference higher than 80th percent of all learning level difference scores as prerequisite concept pairs, the prediction accuracy is 0.39. As shown in Tables 1, incorporating the proposed learning level features does achieve better results than the baseline model. The major reason is that learning level features considers the complexity of a concept, which is a key factor in deciding prerequisite relationships between concepts while the other two sets of features only consider the relatedness between concepts. Moreover, we observe that precalculus have better performance than big data and, when not, is very close. A potential reason is that precalculus is a more fundamental subject than big data and concepts within this domain have more clear learning dependencies.

Table 1: Aaccuracy for mathematics concept map extraction

	40%		60%		80%	
	Baseline	TwoPhase	Baseline	TwoPhase	Baseline	TwoPhase
40%	0.31	0.11	0.46	0.05	0.49	0.16
60%	0.43	0.08	0.61	0.04	0.39	0.1
80%	0.42	0.13	0.55	0.04	0.41	0.08

Table 2: Accuracy for big data concept map extraction

	40%		60%		80%	
	Baseline	TwoPhase	Baseline	TwoPhase	Baseline	TwoPhase
40%	0.35	0.157	0.43	0.114	0.42	0.128
60%	0.39	0.032	0.52	0.1359	0.39	0.154
80%	0.39	0.198	0.51	0.032	0.389	0.098

5. CONCLUSION

We propose the idea of utilizing prerequisite concept map to detect learning gaps for efficient learning. To be specific, we apply a two-phase method for concept map extraction and propose three sets of features for concept relationship identification. Experimental results on two manually created datasets confirm the effectiveness of our investigated features and shows the promise of the proposed model. Moreover, we derive a *top-k* selection algorithm for concepts recommendation.

6. REFERENCES

- [1] R. E. Bennett. Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1):5–25, 2011.
- [2] J. Dermo. e-assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2):203–214, 2009.
- [3] G. Koutrika, L. Liu, and S. Simske. Generating reading orders over document collections. In *ICDE*, pages 507–518. IEEE, 2015.
- [4] C. Liang, Z. Wu, W. Huang, and C. L. Giles. Measuring prerequisite relations among concepts. 2015.
- [5] S. Wang, C. Liang, Z. Wu, K. Williams, B. Pursell, B. Brautigam, S. Saul, H. Williams, K. Bowen, and C. Giles. Concept hierarchy extraction from textbooks. *The ACM Symposium on Document Engineering*, 2015.
- [6] P.-H. Wu, G.-J. Hwang, M. Milrad, H.-R. Ke, and Y.-M. Huang. An innovative concept map approach for improving students' learning performance with an instant feedback mechanism. *British Journal of Educational Technology*, 43(2):217–232, 2012.