# CogLearn: A Cognitive Graph-Oriented Online Learning System

Yang Pian[1], Yu Lu[1 2 *], Penghe Chen[1] and Qinglong Duan[1]
[1]Advanced Innovation Center for Future Educaion, Beijing Normal University, Beijing, China
[2]School of Educational Technology, Beijing Normal University, Beijing, China
*bianyang@bnu.edu.cn, chenpenghe@bnu.edu.cn, duanqinglong@bnu.edu.cn*
[*]*Corresponding Author: Yu Lu, luyu@bnu.edu.cn*

***Abstract*—We propose and implement a novel online learning system, called CogLearn, to support learner's self-awareness and reflective thinking, which urges a proper form of knowledge representation together with individual learner's cognitive status. We thus design and employ the machine learning techniques to estimate learner's cognitive status and identify educational relations to construct the desired knowledge representation, namely *cognitive graph* in our system. We further demonstrate the system by presenting two practical services, i.e., *learning obstacle diagnosis* and *learning path planning*, to demonstrate how the constructed cognitive graph effectively and adaptively supports individual system user's learning process.**

## I. Introduction

As an effective and powerful tool for education, concept map [1] refers to a diagram consisting of instructional concepts and their relations drawn from teacher's learning experience. It has been widely used for teaching and self-learning in schools. Meanwhile, many massive open online course (MOOC) platforms, such as Khan Academy [2], also adopt concept maps to guide the learning process of their users. Cognitive and pedagogical studies have shown that concept map-oriented learning can significantly promote learners' self-awareness and reflective thinking for both online and offline education [3].

However, few existing concept maps used in practice consider individual learner's information, such as their cognitive status, due to the difficulties in both modeling learners and collecting relevant data. Without such key information from the perspective of each learner, concept maps would be hard to support the personalized teaching and adaptive learning in either the offline classroom environment or online MOOC platforms. Moreover, the educational relations between instructional concepts on a concept map are usually labeled by subject teachers or experts. Such a manual construction approach is often error-prone: the pedagogical studies have revealed the so-called *expert blind spot* [4] problem, which means to the same instructional concept, the cognition between experts and learners do not well align in many cases. As a result, even the domain experts or the experienced teachers may easily misunderstand learners' cognitive process and label the improper relations that misguide the learners.

To tackle the above described problems, our insight hinges upon designing a new form of concept map that incorporates individual learner's cognitive information and extracts educational relations by mining learners' data. Specifically, individual learner's cognitive status can be estimated on the concept level (i.e., knowledge status) and properly visualized on the map, by leveraging on the structure information on concept maps with the deep learning techniques. Furthermore, the educational relations are identified by performing the machine learning algorithms on learners' assessment data. Different from the traditional concept maps that simply employ the static and manually labeled relations, this new form of graphic knowledge representation is called *cognitive graph* in our design.

Using the proposed *cognitive graph* as the core component and interactive user interface, we design and implement an online learning system, called CogLearn, to guide user's learning process and provide them personalized learning experiences. Heterogeneous educational data, including the aggregated unit-test and term-test exam data, are collected by the system to support cognitive status estimation and educational relation extraction. It thus makes the following key contributions:

- We propose a novel knowledge representation form and visualization solution, called *cognitive graph*, that properly incorporates the concept-level learners' cognitive status and auto-extracted educational relations.
- We design and implement the cognitive status estimation and educational relation identification models for the construction of cognitive graph.
- We present a practical online learning system, namely CogLearn, and illustrate its personalized learning services directly supported by the constructed cognitive graph.

We will elaborate the cognitive graph and demonstrate the CogLearn system with its key learning services in the following sections.

## II. Cognitive Graph

Briefly speaking, cognitive graph consists of instructional concepts (as nodes in Figure 1) and their interconnected educational relations (as directed lines in Figure 1), typically including prerequisite relation, inclusion relation, causal relation and progressive relation. We automatically extracted the instructional concepts by adopting recurrent neural network models on heterogeneous pedagogical data, such as curriculum standards, textbooks and course tutorials [5]. The cognitive status (i.e., knowledge status) of the current learner is marked on each partially-filled node on the graph, where the number

2375-026X/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDE.2019.00229

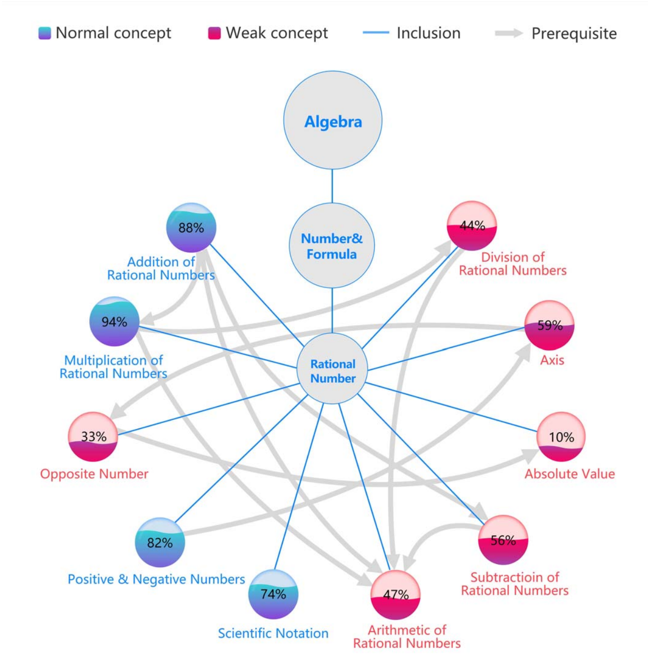



Fig. 1. Exemplary Cognitive Graph for Mathematics

stands for the percentage of mastery on the that concept. To remind the learner of his or her weak concepts, red color is used to highlight the corresponding ones while the concepts with a normal mastery level are usually in blue color. Unlike the traditional static concept maps, cognitive graph provides dynamic information about learners' knowledge status, which can be timely updated using our knowledge status estimation model.

The instructional concepts are mainly classified into four levels in a hierarchical structure, and the concepts on a higher level are represented as a larger size of node. For the subject of mathematics, currently more than 300 fourth-level and 40 third-level instructional concepts are used, and Figure 1 shows all the fourth-level concepts under a third-level concept "rational number", such as "opposite number", "absolute value" and "scientific notation". We see that both the prerequisite and inclusion relations are shown on the graph using blue straight lines and grey arrows respectively. For example, "opposite number" is a *prerequisite* of "absolute value", and rational number includes "opposite number". We will explain the prerequisite relation and how to identify it in the next section.

Based on the above design, we implement the *CogLearn* system to construct the desired cognitive graph in an automatic way and support various upper learning services.

## III. System Architecture and Design

Figure 2 gives the general architecture of the CogLearn system, where the cognitive graph serves as the core component and key interface for learners to access the system. To construct and update the graph, three main modules,

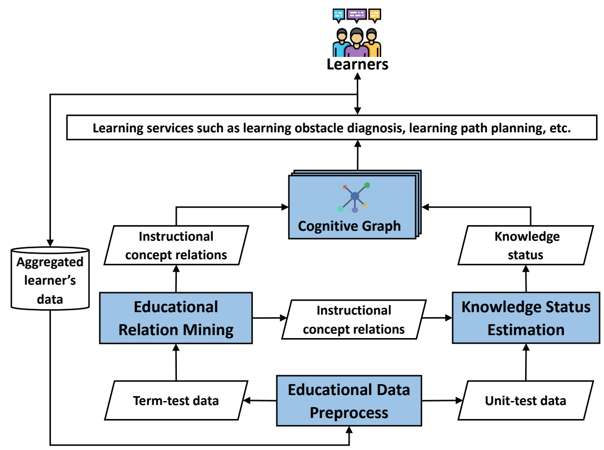



Fig. 2. System Architecture of CogLearn

namely educational relation mining (ERM) module, knowledge status estimation (KSE) module and educational data preprocess (EDP) module, are designed to work cooperatively. The ERM module and KSE module are mainly used for identification of the educational relations and estimation on learner's knowledge status (i.e., the probability of the current learner has mastered that concept) for the cognitive graph. The EDP module mainly preprocesses and provides the required learner's data to other modules, which include user's assessment data and other relevant learning information. A number of learning services have been deployed upon the dynamic and personalized cognitive knowledge graph, such as learning obstacle diagnosis and learning path planning. Learner's interaction data and learning information with such services will be collected and used for timely updating his or her cognitive graph.

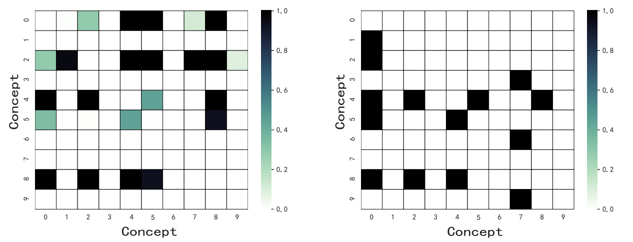
### A. Educational Relation Mining Module

Comparing with other educational relations, prerequisite relation is much more implicit and hard to identify. Thus, we mainly describe how the system identifies it in this subsection.

- **Prerequisite relation identification**: a prerequisite relation from concept $X$ to concept $Y$ means that a learner should master concept $X$ first before proceeding to concept $Y$. Accordingly, given two concepts $X$ and $Y$, if $X$ is a prerequisite of $Y$, learners who do not master $X$ are very unlikely to master $Y$ either. Meanwhile, learners who master concept $Y$ are very likely to have mastered concept $X$ already. It can be formulated as the conditions as follows:

$$M_Y \Rightarrow M_X \quad AND \quad \overline{M_X} \Rightarrow \overline{M_Y} \tag{1}$$

where $M_X$ and $M_Y$ denote learners have mastered concepts $X$ and $Y$ respectively, $\overline{M_X}$ and $\overline{M_Y}$ denote learners have not mastered concepts $X$ and $Y$ yet. Accordingly, the association rule mining [6] can be performed on the learner assessment data to automatically identify such prerequisite relations.




(a) Identified Prerequisite Relations (b) Prerequisite Labeled by Experts

Fig. 3. Prerequisite Relations Among 10 Instructional Concepts (from horizontal to vertical axis).

- **Identification with uncertainties**: as shown in Eqn. (1), prerequisite identification requires the deterministic information on whether the concepts have been mastered or not. However, such information can only be inferred from learners' assessment results, where uncertainties always exist and thus a "probabilistic" approach need to be considered. Specifically, the rules $M_Y \Rightarrow M_X$ in the deterministic association rule will be formulated in probabilistic form $P(M_Y \Rightarrow M_X)$, and it can be held only when the probability value is larger than a given threshold *minprob*. Taking *supp* and *conf*, namely the two parameters of the association rule mining, into consideration, we have:

$$
\begin{aligned}
P\{supp(M_Y \Rightarrow M_X) \geq minsupp \quad AND \\
conf(M_Y \Rightarrow M_X) \geq minconf\} \geq minprob.
\end{aligned} \tag{2}
$$

Considering Eqn. (1) and Eqn. (2) together, determining a prerequisite relation from concept X to Y requires:

$$
P(M_Y \Rightarrow M_X) * P(\overline{M_X} \Rightarrow \overline{M_Y}) \geq minprob \tag{3}
$$

In practice, the ERM module enumerates all the concept pairs in the cognitive graph and calculates the probability of being a prerequisite relation pair using the aggregated learner term-test data. The p-Apriori algorithm [7] can be used to solve the above probabilistic association rule mining problem. This process continues until all the possible pairs are retrieved, and then the identified prerequisite relations are used to construct the cognitive graph. Figure 3 shows an example of the identified prerequisite relations among 10 instructional concepts, where the color in Figure 3(a) indicates the probability calculated from Eqn. (3), and Figure 3(b) shows the results labeled by domain experts.

### B. Knowledge Status Estimation Module

The KSE module mainly employs a deep learning model incorporating the educational relations to estimate learner's knowledge status. Briefly speaking, the module first receives the unit-test data and the identified educational relations from EDP module and ERM module respectively. Subsequently, a recurrent neural network with the GRU unit [8] is constructed and trained to capture the time series information from the aggregated learner's online exercise data. The trained model would be used to update learner's knowledge status on the cognitive graph, where the online exercise data (i.e., the unit-test data) are mainly used as the model input data. By taking the prerequisite relation information into consideration, the proposed deep learning model achieves around 75% overall accuracy and meanwhile can effectively handle the data sparseness problem in learner dataset. Figure 4 shows the designed model, where the prerequisite relations are formulated as the particular constraints in the model, and for more details, please refer to our paper [9].

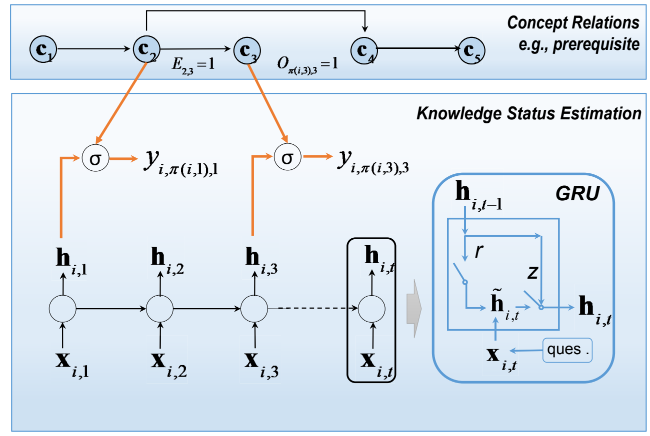



Fig. 4. Knowledge Status Estimation Model using Prerequisite as Constraints

### C. Educational Data Preprocess Module

In the CogLearn system, all the user's learning activities and assessment data would be collected and stored in the system database, which can be accessed from our published repository [10]. The EDP module mainly aims to clean and preprocess such data, and then transfer them to appropriate modules. Briefly speaking, two types of data, namely *unit-test data* and *term-test data*, are mainly used in the system:

- **Unit-test data**: the unit-test is usually associated to one single instructional concept, and normally consists of 9-11 questions. When a learner complete a unit-test, the system will record the information including the question ID, score and the answer submission time, etc. The unit-test data can be directly used for knowledge status estimation.
- **Term-test data**: the term-test data come from the final-term or mid-term examination, which usually contains questions on multiple concepts, and require learner take a few hours to complete. If a learner sits a term test, the system will record the question ID, the corresponding concepts, score and all the other necessary information. The term-test data from a large amount of learners can be used to conduct educational relation identification tasks.

## IV. Services and Demonstration

In our demonstration, we will showcase two learning services supported by the constructed cognitive graph as follows:

- **Learning obstacle diagnosis**: experienced educators often find that a student cannot understand a concept

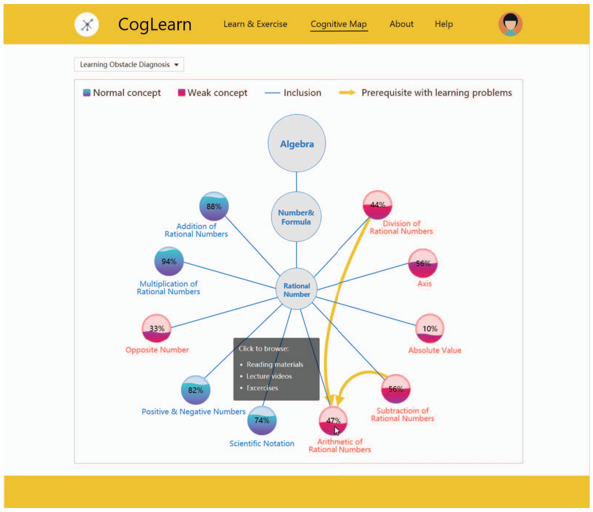



Fig. 5. Learning Obstacle Diagnosis Service

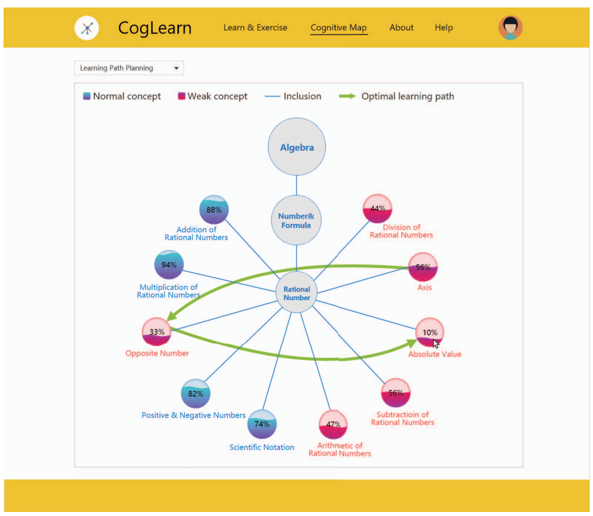



Fig. 6. Learning Path Planning Service

may caused by some relevant concepts rather than itself, especially its prerequisite concepts. Similarly, when a concept is at a low mastery level, the system would automatically check the knowledge status of its prerequisite concepts. Subsequently, the system would highlight all its prerequisite concepts also at a low mastery level, and meanwhile remind that learner using the corresponding learning resources first. As illustrated in Figure 5, the learner clicks the node of a weak concept "Arithmetic of Rational Number", which has four prerequisite concepts in total. Two of them at a low mastery level have been highlighted using the yellow connecting lines, namely the "Subtraction of Rational Number" and the "Division of Rational Number", to help learners aware his or her potential learning obstacles together with the recommendation of learning materials and exercises.

- **Learning path planning**: pedological studies have shown that an optimal learning path would directly increase the learning gain and shorten the learning time. The learning path planning thus enables learners to properly organize their learning schedule and achieve a better academic performance. As shown in Figure 6, when a learner sets up a target concept as "Absolute Value", the system would automatically present the learning "route" step by step on the graph. We see that the current learning path is composed of several concepts that haven't been mastered yet and a heading direction indicates the learning order, that is "Axis", then "Opposite Numbers", and finally the target "Absolute Value". Note that the learning path would be dynamically planned based on the educational relations and the learner's knowledge status.

## V. Conclusion

In this paper, we propose a novel cognitive graph for properly representing and visualizing individual learner's knowledge status and the educational relations on the instructional concept level. We then implement and demonstrate *CogLearn* online learning system, which provides learning obstacle diagnosis and learning path planning based on the constructed cognitive graph. We are working on deploying the CogLearn system to serve more than 4000 local students.

## Acknowledgement

This research is supported by the National Natural Science Foundation of China (No.61702039, 61807003), the Humanities and Social Sciences Foundation of the Ministry of Education of China (No. 17YJCZH116) and the Fundamental Research Funds for the Central Universities.

## References

[1] A. J. Cañas and J. D. Novak, "The theory underlying concept maps and how to construct and use them," *Práxis Educativa*, vol. 5, no. 1, pp. 9–29, 2010.

[2] K. Academy, "Khan academy," https://www.khanacademy.org/, accessed November 29, 2018.

[3] G. Y.-M. Kao *et al.*, "Breaking concept boundaries to enhance creative potential: Using integrated concept maps for conceptual self-awareness," *Computers & Education*, vol. 51, no. 4, pp. 1718–1728, 2008.

[4] M. J. Nathan and A. Petrosino, "Expert blind spot among preservice teachers," *American Educational Research Journal*, vol. 40, no. 4, pp. 905–928, 2003.

[5] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, "Knowedu: A system to construct knowledge graph for education," *IEEE Access*, vol. 6, pp. 31 553–31 563, 2018.

[6] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *ACM Sigmod Record*, vol. 22, no. 2. ACM, 1993, pp. 207–216.

[7] L. Sun, R. Cheng, D. W. Cheung, and J. Cheng, "Mining uncertain data with probabilistic guarantees," in *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2010, pp. 273–282.

[8] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[9] P. Chen, Y. Lu, V. W. Zheng, and Y. Pian, "Prerequisite-driven deep knowledge tracing," in *Proceeding of the IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018.

[10] B. N. University, "Smart learning partner dataset," http://www.bnu-ai.cn/data, accessed November 25, 2018.