Develop Academic Question Recommender Based on Bayesian Network for Personalizing Student’s Practice

Qingsheng Zhang (*), Di Yang, Pengjun Fang, Nannan Liu, Lu Zhang
Xi'an University of Posts and Telecommunications, Shaanxi, China
zhangqs@xupt.edu.cn

Abstract—Study in Literatures shows that tracing knowledge state of student is corner stone of intelligent tutoring system for personalized learning. In this paper, an academic question recommender based on Bayesian network is developed for personalizing practice question sequence with tracing mastery level of student on knowledge components. This question recommender is discussed with theoretical analysis, and designed and implemented in software engineering way. It provides instructor with tools for building knowledge component network and setting question of course. It also makes student personalize practice questions of course. This question recommender is planned to deploy in real learning context for the future validation of how well such question recommendation improves performance and saves practice time for student.

Keywords—Question recommender, student academic capability, knowledge component, Bayesian network, software engineering.

1 Introduction

Information technology enhanced education makes the continuous improvement of teaching and learning over years. For example, a cognitive conversational agent based on domain ontology is implemented and used to optimize tutoring interaction with students in their self-study processes in a kind of positive emotional context, in which a well-defined set of short talk segments operated by this cognitive agent is in the form of questions and answers interaction as stated in [1]. In another example, intelligent evaluation methodology based on fuzzy logic and knowledge based expert system are developed for the automated evaluation of student’s learning and providing teachers with indicators on the students’ strength about their learning difficulties, false studied knowledge components, which also can be used to regulate the learning process as stated in [2]. Such traditional intelligent tutoring system and adaptive e-learning system are familiar to educational field.

Things always change, as item recommendation based on users’ similarity in e-commerce field, recommender system is also applied to education, and now educational recommender system is one of focuses in educational technology field. For
example, a recommender system based on statistical cluster techniques are developed for the recommendation of relevant learning resources and like-minded student peers as stated in [3]. For traditional e-learning system, more effectively use and combine the recommendation of student peers and learning contents are discussed for the enhancement of learning outcome as stated in [4]. In the context of Massive Open Online Courses (MOOCs), questions recommendation in discussion forum with context-aware matrix factorization model based on the constraints of both student’s answering question capacity and expertise matching of requested question is studied for improving predicting student’s preferences over questions as stated in [5]. Recommending courses to online students in MOOCs with hybrid weight course recommendation algorithm based on the latent interest model, user demographics and course prerequisite is studied for improving students’ course enrollment behavior as stated in [6]. A concept-based recommender system for recommending learning materials is proposed to meet the pedagogical goals of instructors during the creation of an online programming course, and decrease the time that the instructors spend on authoring task and keep the coherence of the sequential structure of the course as stated in [7].

In above these educational recommender systems, the student peers and courses as item recommendation are similar to those e-commerce recommender systems termed as neighborhood-based. To master detail knowledge components in a course, students prefer a fine grain items (that is, practice questions on knowledge components in a course) recommended to student peer or course recommended. In such educational recommender, question recommendation is based on the probability prediction of mastering knowledge component by using Bayesian networks [3, 8] rather than neighborhood-based recommendation. However, such literatures are rarely presented in this aspect of educational recommender system. The proposed academic question recommender is considered as students practice proper question opportunities in a personalized way based on their knowledge mastery level. Skill development and expertise of human are strongly related to the time and efficiency of deliberate practice as stated in [17]. It means that more practice with additional time spent yields improved performance. However, how much practice should be included in e-learning course depends on the criticality of proficiency expected by instructor and student. The statistics as stated in [18] show a minimum of seven practice opportunities is needed for each knowledge component in their studied 74 courses in four topic domains. However, the difference of knowledge level, background and cognitive style among students is not included in the practice opportunity study. Personalized learning path also helps students fill in the gaps in their knowledge, improve their performance in curriculum level and project level as stated in [19,20]. To solve the issue of individual student’s proper practice opportunities for mastering knowledge in individualized practicing questions of knowledge components in e-learning course, a question recommender based on Bayesian network is proposed.

The rest of paper is organized as follows. In section 2, literature review about both recommender system and intelligent tutoring system based on Bayesian network are presented. In section 3, the method of recommendation and the technique of software engineering are separately discussed for the question recommender. The key functions
of question recommender are snapped for explanation in section 4. The paper is concluded with a discussion of question recommender and future work in section 5.

2 Literature Review

The prediction and recommendation task of recommender system is considered as estimating the expected value of user’s rating on unseen item according to a probabilistic perspective. Therefore, Bayesian network for probabilistic uncertainty is used in recommender system. The authors as stated in [3] merely mentioned Bayesian network as possible recommending method without any extended discussion. Recommender system based on Bayesian network as stated in [8] is fully described in theoretical analysis as a node of Bayesian network corresponding to each item in the domain. The states of each node correspond to the possible rated values for each item. Resulting Bayesian network, in which each item may have a couple of parent items for indicator of best predicted its rating, is obtained by applying learning algorithm to training data. Case studies show that Bayesian network is used to predict and recommend in concrete applications. For example, an insurance recommender system using Bayesian network as stated in [9] predicts better than traditional low rank matrix factorization model for insurance domain. Next, an overview of intelligent tutoring system based on Bayesian network is discussed.

The possible factors of modeling student are learning context, preference, interest, goal, motivation, cognitive style and knowledge state et al. Among them, tracing knowledge state of student is most important and successful student modeling technique. Tracing knowledge state of student as stated in [10] is a study of modeling students’ changing knowledge state during skill acquisition in practice environment for students’ writing codes in programming language. To each individual knowledge component, an initial learning probability is given assumed that it is in the learned state prior to the first opportunity to apply this knowledge component, and an acquisition probability is also determined for the transition from the unlearned state to the learning state following an opportunity to apply knowledge component. This tracing knowledge state model also has the other two probabilities about student guessing correctly in the unlearned state of knowledge component and student slipping wrongly in the learned state of knowledge component. Therefore, the state of knowledge component is estimated with the sum of posterior probability of knowledge component already in the learned state as evidence and probability of transition from the unlearned state to the learned state after a new opportunity of practice, and then this tracing model makes prediction at each step depending on the student’s knowledge state. Bayesian inference scheme is used to estimate posterior probability in this tracing knowledge state model.

A theoretical sound works as stated in literature [11] is proposed to use Bayesian network to structure knowledge components of course in prerequisite relations. It also provides two tasks of intelligent tutoring system: gathering information of student’s current state on knowledge component and modeling students’ unreliable source of information, such as students’ guessing, slipping etc.
As stated in literature [12], an intelligent tutoring system presents the lecture notes, examples and quizzes of course. This tutoring system employs Bayesian network to make decision of how to support students in learning basic programming language. In other words, mastery level on knowledge component in conditional probability table of Bayesian network is obtained with the result of previous course final exams, then a student reads lecture or answers quiz, and this system decides if a basic knowledge component is understood by student. The conditional probability tables in Bayesian network are updated, and the mastery level on knowledge component is evaluated again. It presents prerequisite knowledge components if the current knowledge component is not understood. It also generates a proper learning sequence for the all unknown ancestral knowledge components of student. However, the authors do not report the study of effect on tutoring system in students’ real learning context.

A conclusion is that knowledge state tracing system models the student’s state change on knowledge components from an ideal student model with practice opportunities and presents an individualized sequence of exercises to the student; however, educational recommender system calculates the mastery level of student on knowledge component in the Bayesian network by human domain expert. Individualized and self-controlled practice opportunities on knowledge components is not studied in above mentioned both systems where only a fixed mastery probability is provided. The proposed academic question recommendation system tries to provide the solution of optimal practice opportunities in knowledge component network with an individualized mechanism which is studied and developed in literature [16]. Currently, a basic version of question recommender is deployed in student real learning environment for practicing questions on knowledge components from computer science and software course. It is also gathering data about student’s action and performance for validation of optimal mechanism on practice opportunities in next step research.

3 Method and Technique

Bayesian network is applied to the proposed question recommender with tracing mastery level of student on knowledge component and structure of knowledge components in course for personalizing practice of student. The question recommender consists of the module of student answer, the module of student’s academic capacity evaluation, the module of instructor building knowledge component network, the module of instructor setting question and the module of Bayesian inference module.

In learning context of question recommender, an instructor builds knowledge components, sets questions and determines Bayesian network for a course. A student finishes question answer on all knowledge components in the course. Then the evaluation module assesses question and calculates mastery level of student on each individual knowledge component. Bayesian inference module predicts the probability of unmastered knowledge components with the academic capacity of student and the Bayesian network of course and recommends the questions in descending order of probabilities of the unmastered knowledge components for the student, who practices the questions in a personalized sequence.
3.1 Bayesian network for question recommendation

Bayesian network is represented with Directed Acyclic Graph (DAG), in which nodes represent a group of random variables and edges represent the dependency relation between the random variables. An example of Bayesian network is shown in Fig. 1 as follows. Node A is parent of node B and C, and D is also B’s another parent. Node E’s direct parent is B. Node B is conditional independence on C, which is not B’s child. That is, B conditionally independent of C that is not descendent of B.

![Fig. 1. Bayesian Network Example](http://www.i-jet.org)

To calculate the value of joint probabilities given by a Bayesian network, each node in Bayesian network should be with probability. The probability of nodes without parents is called prior probability, such as node A and D in Fig. 1. The rest’s probability is conditional probability on their parent nodes, such as node B, C and E in Fig. 1. Formally, Bayesian network is a pair \((G, T_p)\), where \(G\) consists of a set of nodes and directed arcs with the given conditional independence assumptions, and \(T_p\) is a set of conditional probability distribution table associated with each node in the network. Let \(g_1, g_2, \ldots, g_n\) be the nodes in a Bayesian network. Given the conditional independence assumptions made by this network, a joint probability of all of nodes in the network as follows:

\[
P(g_1, g_2, \ldots, g_n) = \prod_{i=1}^{n} P(g_i|\text{Parents}(g_i))
\]

(1)

where Parents \((g_i)\) is a set of direct ancestor nodes of node \(g_i\).

Two patterns of inference in Bayesian network are used by question recommender. One of them is causal reasoning, that is, the evidence of some nodes in Bayesian network is available, and their descendant node is inferred with the evidence. For exam-
ple, node B has two parents A and D in Fig. 1. Descendant B’s conditional probability is expanded into the sum of joint probabilities of B and all parents as follows:

$$P(B|D) = P(B \cap (A \cup \neg A)|D) = P(B, A|D) + P(B, \neg A|D)$$

(2)

This joint probability is back to the probability of B conditioned on all of the parents. That is, Equation 3 is obtained by a form of condition probability as follows:

$$P(B|D) = \frac{P(B, A|D)p(A)}{P(D)} + \frac{P(B, \neg A|D)p(\neg A)}{P(D)}$$

(3)

Equation 3 is reduced to Equation 4 as A and D in Fig. 1 is independent.

$$P(B|D) = P(B|A, D)p(A) + P(B|\neg A, D)p(\neg A)$$

(4)

Another inference in Bayesian network is diagnostic reasoning. That is, the symptom is used to analyze a possible cause, which is the application of Bayesian probability rule. The key step is the use of Bayesian probability rule to convert the analysis problem into one of causal reasoning with Equation 5 as follows:

$$P(\neg D|\neg B) = p(\neg B|\neg D) \times p(\neg D)/p(\neg B)$$

(5)

State value in Equation (6) is an indicator for mastery level of student on knowledge component. The criterion of mastering knowledge component is 0.95 as stated in [10]. However, it is also set to 0.8 in some literature, so this criterion is determined by the teaching experience and domain expertise of instructor. Correct and incorrect in Equation (6) represent the numbers of right answering question and wrong answering question on knowledge component. In other words, frequency of counting answer on knowledge component is considered as probability of node in Bayesian network.

$$state\ value = \frac{correct}{correct + incorrect}$$

(6)

Two patterns of inference in Bayesian network employed by question recommender is a kind of formal framework for probabilistic reasoning in uncertain context. Casual reasoning can recommend questions on the unmastered knowledge components in the descending order of probability. A student goes through a personalized learning path under the unmastered knowledge components and their ancestor. Thus, the nodes in Bayesian network are knowledge components in a course, and the edges between nodes reflects the prerequisite dependencies that hold among knowledge components, or else conditional independence holds if the directed edges do not exist among knowledge components.

### 3.2 Workflow of question recommender

The workflow of question recommender is shown in Fig. 2 as follows. First of all, a domain expert, such as math instructor, needs to propose knowledge components
and build the corresponding network Bayesian with them for the course. The domain expert also sets questions on each individual knowledge component in the course. Then, the student can practice the questions on all knowledge components, whose answers are assessed and the results are recorded for the analysis of student academic capacity in the course.

Under the knowledge component network of course and assessment of student’s practice, casual reasoning of Bayesian network calculates the probability of descendant node with the probabilities of their parents. The domain expert stipulates the criterion of mastering knowledge component. The knowledge component is not mastered by the student if the correct rate of student on some knowledge component is lower than the criterion. By the casual reasoning in knowledge component network, those knowledge components with correct rate lower than the criterion are selected and sorted in descending order. The question recommender personalizes the practice of students and easily improves the performance of students on assumption that is first recommending questions on knowledge component with higher correct rate for the unmastered knowledge components.

3.3 Develop question recommender

The question recommender is developed in the way of software engineering project. To develop this recommender, unified modeling language (UML) [13] is used in the requirement analysis and design of question recommender. First step in development of question recommender project is requirement analysis. The key requirements of question recommender are shown with use case diagram of UML in Fig. 3. The first scenario in question recommender is that domain expert selects knowledge components from a course and determines their relations. Then questions on each individual knowledge component are prepared for the student’s practice on knowledge components. The second scenario is that the student initially practices the questions on all
knowledge components of course, the recommended questions are present to the student by casual reasoning of question recommender, and then goes through a personalized question sequence.

Fig. 3. Use Case of Question Recommender

Besides product use case, other products in requirement analysis are domain model and system sequence diagram which are also components of UML. Such products with the different views of requirement analysis are not presented in this paper for its research purpose. Next step in development of question recommender project is software design based on the products of requirement analysis. Like requirement analysis of software project, the detailed and full design information is necessary for the project development. However, it is not essential for the purpose of academic and technical research. To clearly understand the question recommender, some key design products of recommender are described with class diagram and sequence diagram of UML. In Fig. 4, class diagram of knowledge component network is provided for understanding basic characteristics and operations in question recommender. The interface named "NodeService" is implemented by the class "NodeServiceImpl", which is also dependent on others interfaces and classes in Fig. 4, for creating node, querying node, deleting node, revising node name, making knowledge component network and updating parameters of Bayesian network. Term "JSONObject" in Fig. 4 originates from JavaScript Object Notation [14] that is a lightweight data interchange format between student client console and back end server of question recommender. Again, as the research purpose of this paper, each individual attribute and operation of classes
are not discussed although they are concretely designed in the software project of question recommender.

**Fig. 4.** Class Diagram of Knowledge Component Network

Information in Fig. 4 is about what components of question recommender contribute to the building Bayesian network, however, how interaction and what sequence between these components are not available in Fig. 4. The interactive information and sequence are also important for software design of question recommender.

In Fig. 5, abovementioned sequence diagram is used to describe how to interact between instances of classes in creating knowledge component network. Rectangle with specified name in Fig. 5 is symbol of class instance for the participants of creating knowledge component network. Vertical broken line is time line for classes’ instances during their life. Vertical thin rectangle represents the activation state of corresponding instance. Horizontal line with arrow represents the calling function or sending message between classes’ instances. With available information in Fig. 5, the interactive process among different classes’ instances and the function of task are clear to understand for readers.
In Fig. 6, the process of question recommendation is described with sequence diagram. In this sequence diagram, the interactive information is printed above the lines with arrows, the sequence of interaction is in order of number, and the required participant instances of recommending question for student are also presented for understanding the how of completing question recommending task. To compare with class diagram, sequence diagram shows more detailed and full interactive information for software coding.

Another important task of question recommender is assessment of question on knowledge component. Classes’ instances and interactive information in sequence diagram of question assessment are presented in Fig. 7 as follows.
Software design products in Fig. 4, 5, 6, 7 make the designer and programmer work and coordinate in the same idea for the development of question recommender. The last step in development of software project is software coding and testing of question recommender, which is fulfilled based on all software design products. Any coding snippet for implementation of question recommender is not mentioned in this paper as theoretical analysis and software developing technical details provided in this section is available for understanding recommender. Key snapshots of question recommender in run time are shown for the demonstration of question recommender in section 4.

4 Key Views of Question Recommender

The snapshot of student question practice is presented in Fig. 8. The questions on all knowledge components in a course selected by student are shown in the initial practice. The left side grey area in Fig. 8 is the navigation of knowledge components in course. The questions on knowledge component are present in upper right practice area in Fig. 8 when the student clicks one knowledge component. The blue button in the right side of practice area is used to present next question. Seeing hint in red in left side of practice area is used to provide possible help message on the question. The function of question choice under practice area allows student to pick one of questions on a knowledge component for answer. The answer is presented to the student after all questions on knowledge component are finished and submitted. The recommended questions are present if the student finishes questions of all knowledge components of course. In other words, the red button “recommended question” in the bottom of
knowledge component navigation area is activated, the student clicks this red button and practices the recommended questions of unmastered knowledge components.

Two samples of recommended question sequence on the unmastered knowledge components are shown in Fig.9 for a simple trial of personalized practice, which is snapped in the console of back end server. The knowledge component node identity and name of recommended question are circled in red and blue separately. Two samples are noticed to the difference in the node identity sequence for the demonstration of students personalized practice function being provided by question recommender.

An instructor uses the available tools to set questions on knowledge components of course and build the corresponding Bayesian network by identifying relations between the knowledge components. In Fig.10, it is a snapshot for the management of question setting. The bottom in navigation area in left side is the question management with red ellipse. Question setting area in right side is present while the instructor
clicks question management in navigation area. Draw down list box in red ellipse in top left corner in question setting area is used to choose one of available courses, then question list table of selected course is presented in the bottom of question setting area. By clicking adding button in red ellipse or modifying button not shown in the right side of question setting area only with right arrow, the instructor can add or update question content and corresponding knowledge component. Other parts in red ellipse in Fig. 10 are shown with comment, the discussion are not given again.

Fig. 10. Question and Knowledge Component Management

In Fig. 11, the big red ellipse in navigation area in left side is building Bayesian network, by clicking on which the management of knowledge component network is presented in right side. The drop-down list box in green ellipse in the right side is for selecting one of available courses. Once one course is selected by the instructor, all available knowledge components of course are shown for building knowledge component network. There are some red circles in the middle part of right side. These red circles represent knowledge components of course.

There are also buttons of basic operation alongside the drop-down list box in the top of right side in Fig.11: displaying relation between nodes, adding new relation between nodes, verifying the resulting network DAG. Clicking on the button of adding new relation between nodes, two columns of drop-down list box are presented in the bottom of right side, which are used to determine relation between ancestral node and descendant node. By clicking on the button of verifying the built knowledge component network DAG, the information of verified knowledge component network is sent to the back-end server as Bayesian network. Another small red ellipse in navigation area is node management for adding new knowledge component of course. Then the new added knowledge component is used in the task of question management or building Bayesian network.
Discussion and Conclusion

The implementation of question recommender is accompanied by software testing of key functions. This question recommender provides instructors with a platform for creating knowledge component network in their courses. It also helps students practice their unmastered knowledge components in personalized sequence. However, it is not sure this recommender really improves the performance of student and saves practice time without deploying into a real learning context for collecting data and analyzing its effects. It plans to deploy this recommender into real learning context for this purpose. In the current version of recommender, only casual reasoning of Bayesian network is available, diagnostic reasoning is not implemented for question recommendation.

Additionally, question and knowledge is one to one mapping in current version of questions recommender. However, it is not true for setting question in one to one mapping as one question mapping to more than one knowledge components is possible [15]. Relation between knowledge components is traced with Bayesian network in this paper for casual reasoning. However, learning curve on more practice opportunities on a knowledge component also contributes to the prediction and recommendation. The study results from learning curve as stated in [10, 16] is integrated into this recommender for optimizing the process of recommending question as decision factor.

In this paper, a question recommender based on Bayesian network with students’ academic capacity is developed for the personalized practicing question sequence. Next, the diagnostic reasoning function of Bayesian network will be developed. This
question recommender is prepared to deploy in real learning context for the evaluation of its effects on students.

6 Acknowledgement

This work is supported by Higher Education Teaching Reform Research Project of Shaanxi Province (No. 19BY091).

7 References


8 Authors

Qingsheng Zhang is with Xi’an University of Posts and Telecommunications, Xi’an, China. His research interests include ubiquitous computing, context-aware computing applications, distributed computing and applications, data mining, and adaptive eLearning. (email: zhangqs@xupt.edu.cn).

Di Yang is now with Alibaba Group. (email: Dick_YangDi@163.com).

Pengjun Fang is now with Kingdee Apusic Cloud Computing Co. Ltd. (email: fangpj1981@163.com).

Nannan Liu is now with Archforce Financial Technology. (email: lakesi_nan@163.com).

Lu Zhang is now with Alibaba Group. (email: luis404@163.com).