Toward an Adaptive Learning System Framework: Using Bayesian Network to Manage Learner Model

Viet Anh NGUYEN

University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam

Abstract—**This paper represents a new approach to manage learner modeling in an adaptive learning system framework. It considers developing the basic components of an adaptive learning system such as the learner model, the course content model and the adaptation engine. We use the overlay model and Bayesian network to evaluate learners' knowledge. In addition, we also propose a new content modeling method as well as adaptation engine to generate adaptive course based on learner's knowledge. Based on this approach, we developed an adaptive learning system named is ACGS-II, that teaches students how to design an Entity Relationship model in a database system course. Empirical testing results for students who used the application indicate that our proposed model is very helpful as guidelines to develop adaptive learning system to meet learners' demands.**

Index Terms— **Adaptive Hypermedia, Learner Model, Bayesian Network, ACGS-II.**

I. INTRODUCTION

In the last decade, with the innovation of internet technology, web-based training systems have become increasingly popular in education. However, hardly do the learners obtain knowledge that they need because of huge course content information. Adaptive educational hypermedia systems (AEHS) are designed to develop the courses that their content can adapt to users' demands. There are many AEHSs which are developed to meet learner's demands such as AHA [1], KnowledgeTree [2] and KBS Hyperbook [3]. These system focus one or more learner's information such as knowledge, background, learning goals, preferences can be used as factors for adaptation.

Up to now, there are many approaches to develop a learner model such as topic-based learner modeling, concept-based learner modeling, generalized domain model, generalized overlay model [4]. Each approach aims to focus on one or more useful features of learner information and it has some benefits and shortcomings. For example, the topic-based learner modeling is easier for learner and teacher to grasp, to index content and to clear interface for presentation but the learner model of this approach is too coarse-grained and precision of learner modeling is low. The concept-based learner modeling overcomes the shortcomings of topic-based learner modeling but it also confronts with top-based learning modeling benefit. How to construct and manage learner model in order to efficient adapt process remains a

challenging question for researchers. Overlay knowledge modeling, which is the most popular for AEHS, presents an individual learner's knowledge as a subset of domain model. Learner model needs to store data to estimate learner knowledge level about concept, which is a part of domain model. Therefore, it is not easy to precisely evaluate learner knowledge level. Using Bayesian Network (BN) to manage uncertain factor in overlay model is a good approach. Because it can be used for various domains and its structure-resemble knowledge network model in which concepts were connected by different kinds of relationships, can be represented by parameters [5]. In CATs system [6], BN was used in order to select new question for adaptive test, the network was constructed as node that measures student's knowledge and gathers evidence with two kind of links: aggregation relationships among knowledge variables, and relationships among knowledge and evidential variables.. SQL –Tutor [7] presents domain knowledge in term of many constraints, which are factors for BN to make multiple predications about student performance. In our previous work, ACGS [8] system also used BN to generate a learning path based on learner's learning goals.

About the course content modeling, there are several systems representing the course content as a set of concepts [9][10]. However, learners not only need to learn concepts but also need to have skills to do several tasks, which was required as part of the course domain. One of the course objectives is that the learners need to apply learned concepts to make these tasks. In addition, one of the difficulties that learners may face is how to learn the concepts by a recommended adaptive system. From our point of view, the disadvantage of these systems [2,3,11,13] is that they only recommend concepts which learners need to learn but did not give more instructions how to acquire the concepts. Objective of our research is to develop basic components of AEHS framework in order to adapt course content as well as to recommend learners how to acquire the suggested concepts. Therefore, we promote a new approach to represent the learner model, the domain model and the adaptation engine.

The paper is organized as follows. The next section represents an overview of theoretical background. It starts with a description of several main approaches to adapt and briefly describes components required for overlay knowledge modeling: the domain model, the overlay knowledge model, as well as BN for constructing and managing a learner model. The third section represents our theoretical studies for developing a learner modeling,

a domain model. These are base not only for developing the mechanism of choosing learning tasks in accordance with the knowledge of each learner, but also for choosing the suitable learning process with the objectives of the studies. The fourth section is devoted to our Adaptive Course Generation System (ACGS-II). It describes the system architecture and how to apply new approach to manage the learner modeling, the domain model and the adaptation engine in its. The next section focuses on analyzing the empirical results of ACGS-II for students who are participating in a computer course. Discussion and conclusion of our work are represented in the following sections.

II. THEORETICAL BACKGROUND

This section describes several theoretical backgrounds, which involved in our research. This presents some main approaches to adapt learning materials and briefly describes components required for a overlay knowledge modeling such as the domain model, the overlay knowledge model.

A. The domain model

Domain model is an object model of problem domain. In AEHS, domain model is a set of elements of educational domain; each element is a domain object class and the relationship between them. Domain model decomposes knowledge of the subject into fragments such as topic, sub-topic, and atomic concepts. Depending on the domain, there are many kinds of domain model structure: vector model, network model, and ontology, etc. Let see [4] for more details. In this paper, we only focus on network model, which is used to construct domain model of ACGS-II. In network model, several links that represent different kind of relationships between concepts. Aggregation and part-of relationships are popular kinds of links that were used in many systems. In the former, mastering all sub-topic causes mastering topic, in the later, mastering topic causes mastering all sub-topic. KBS system [14] , INSPIRE [15], and NetCoach [16] are some of frameworks using aggregation or prerequisite relationship for domain model. Meanwhile, AHA! [1], MEDIA [5], and DCG + DTE [17] are some of frameworks using part-of relationships for domain model.

B. Overlay learner modeling

The overlay model supposes that the student's knowledge is a subset of the system knowledge of the subject. As the student learns, the subset grows, and the modeler's job is to keep track of the subset. This model assumes that the student will not learn anything that the expert does not know. The principle of the learner's overlay model is that for each domain model concepts, individual users' knowledge model stores data that represent values, which is an estimation of the user knowledge level of this concept. This estimation can have discrete value [1,19], which uses a quantitative value to represent the level of learner knowledge, or probabilistic [10,20] values, which use the form of uncertainty management such as fuzzy logics or BN to manage learner knowledge. Therefore, in practice, overlay models of individual learners stored a set of name-value pairs in which the name indicates domain model concepts and the value denotes the level of learner's knowledge.

C. The Task Model

A task statement refers to a set of coherent activities that are performed to achieve a goal in a given domain. The mechanism of hierarchical and recursive decomposition of a problem into sub-problems is one of the basic characteristics of the hierarchical task model [21]. There are several properties for tasks such as: A task statement describes a finite independent part of the job, a task statement uses one verb, and must be measurable, etc. The task model, which stores the results of the task analysis process, is used to find out several activities of people and to establish requirements for training and for user documentation. Task models are documentation structures that are used for: i) documenting the result of a task design of proposed activities, ii) supporting personnel selection, iii) identifying needs for training.

III. DOMAIN AND LEARNER MODELING

This section focuses on our new approach for representing the course content model and the learner model. It discusses the elements of course content model, their attributes and their relationships. It also describes how to represent learner model in order to adapt the course content based on learner's knowledge.

A. The course content model

Course content usually includes concepts, so we use the concept to model content. This trend is consistent with current researches in AEHS as there are many models [11] [9] [10]which also used the concept as one of the element of the course content. The difference in the usage of concepts in these studies is to determine the unit to measure a concept. Depending on the domain, different applications and perspectives of the design measure different the concept such as knowledge [22] , rules [23], and constraints [7]. In addition to the knowledge that learners need to learn, the course content also includes the tasks that learner need to finish in order to achieve the goals of the course. Therefore, the domain model should include the task.

1) Elements of the course content

In our model, we propose concepts and tasks as components of the course content by following definitions:

Definition 1: A concept is a basic unit to present a specific content.

In the content model, the concept is understood as the smallest unit of course content, in other words, it would not exist as a C_i concept, which is a part of C_j concept. To determine the relationship between these concepts, we propose the prerequisite concept definition as follows.

Definition 2: Prerequisite concepts: C_i is called a concept's prerequisite of C_j concepts in order to understand the C_j concepts necessary to understand the concept C_i (Denotes: $C_i \rightarrow C_j$).

Defining prerequisite described relations between concepts in the model, we only consider the prerequisite relationships between concepts rather than considering the relationship component which is used in some other models [1,5], because composition concept is considered as the smallest unit in our model. The conceptual model is illustrated as a graph, in which vertices of the graph are the concepts, the edge shows the relationship between

concepts. To show information of the learners' knowledge level, we developed a model based on the overlay model. The overlay model allows evaluating the level of learners' understanding with all objects in the domain. In addition, in the content model study, because of the relationship between required objects, assessment of learners' understanding of concepts is considered in their relationship instead of considering the independent objects. That is why that we selected this model to represent information knowledge level of learners. We use the probability value to quantify the level of understanding for the learner's concept because of following reasons: i) the understanding level of the learner through the results of test questions and exercises is an uncertain factor; there is no absolute precision in determining the value of understanding level of each concept. ii) Qualitative value (good, average, poor, etc…) or quantitative values ([0...100]) is not precision in related quantitative concepts. iii) Using network model to represent course content for examining the concept relationships.

Definition 3: Task is unit of work that learners need to finish in the learning process in order to fulfill the course objectives.

Unlike the concept, the task requires learners to interact with the system. Unlike the Choquet promoted (1998), which consider learning concept is a task, we define the task as homework or learning activities that require learners to apply acquired concepts to solve. To determine the relationship between tasks, we propose definition of prerequisite task and component task as follows.

Definition 4: Prerequisite task: T_i is called the prerequisite of T_j task if to finish the task T_j, the learner must finish the T_i task (symbol $T_i \rightarrow T_j$)

Definition 5: Component task: T task includes T_1, T_2, …, T_n with $T_1 \cup T_2 \cup ... T_n = T$ and T_i, T_j $(i \neq j)$ $T_i \cap T_j = \emptyset$. T_i (i=1..n) is called component of T task.

B. The learner modeling

The learner model is an important component for building adaptive course as well as for the basis of classification of learner to evaluate and build the corresponding learning content for each learner [25,26]. The learner model includes assumptions, information to represent characteristics of learners. In this study, we develop learner modeling through a new approach to manage learner's knowledge, learner's preferences. We describe the knowledge level of learners using state variables and probability values to quantify the level of understanding of learners, using probabilistic BN model to quantify the level of knowledge learners with the related concepts and tasks. The quantitative value of knowledge is a basis for suggesting peoples need to learn concepts how in order to complete a task.

We also supply some properties to represent learner's demand and objectives, which are fundamental to create the learning path for many learning goals of learner instead of recommending it to meet individual goals, individual needs, such as the approach of some models did [10,13,27,28].

1) Modeling learner's knowledge

In the model, with each concept we use two state variables to quantify the level of knowledge of learners because of following reasons: i) the overlay model needs a variable to store value indicating level of knowledge to the

learner's concept. ii) Assess the level of learners' knowledge is needed for quantify concepts. In the model, we represent each level through a state.

For each C concept, the two state variables are used to measure their understanding of learners. It is:

- *Not_acquired:* represents the level of learners' knowledge that does not acquire the concept.
- *Acquired:* represents the level of learners' knowledge that acquires the concept.

For each C concept, p (C = not_acquired), p (C = acquired) denotes the probability value representing the state may be not acquired or acquired the C concept. It has: p(C = not_acquired) + p(C = acquired) = 1.

Bayesian Network was used to quantify the level of understanding for the learner's concept because of several following reasons: i) Course content is modeled by the network model, considering the concept of objects with interdependence. The concepts and their relationships in the content model establish causal Bayesian probability network. ii) Probability value is used to quantify the level of learners' knowledge to the concept. iii) Considering relations $C_i \rightarrow C_j$, BN are probabilistic reasoning mechanism and diagnostic reasoning mechanism which help to predict the level of understanding of the C_i concept when has known the level of understanding quantitative C_j concept and reverse.

From the general formula for the probability distribution, we determined the quantitative formula of the level of knowledge for C_n concept through the following propositions:

The C_1, C_2, …, C_{n-1} concepts are the prerequisite concepts of C_n concept. Meanwhile the value of quantitative level of knowledge of learners to the C_n concept is determined by the following formula:

$$P(C_n| C_{n-1},...,C_2,C_1) = P(C_n| P_a(C_n)) \text{ with } P_a(C_n) \subseteq \{ C_{n-1},...,C_2,C_1\} \qquad (1)$$

Prove:

By definition of the conditional probability, we have:

$$P(C_1, \ldots, C_n) = P(C_n|C_{n-1}, \ldots, C_1) * P(C_{n-1}, \ldots, C_1)$$

Continue to implement the formula we obtain:

$$P(C_1, \ldots, C_n) = P(C_n|C_{n-1}, \ldots, C_1) * P(C_{n-1}|C_{n-2}, \ldots,$$

$$C_1) * \ldots * P(C_2|C_1) * P(C_1)$$

$$= \prod_{i=1}^{n} P(C_i | C_{i-1},...,C_1)$$

From the general probability formula and Cn variable depends only on the parent node of the set $P_a(C_n)$, we obtain $P(C_n | C_{n-1}, ..., C_1) = P(C_n | P_a (C_n))$.

Similarly, for each T task, the two state variables are used to measure completing of learners to this. It is:

- *Not_finished:* represents level of learner that does not acquire the task
- *Finished:* represents level of learner that acquires the task.

For each T task, p (T = not_finished), p (T = finished) denotes the probability value measure the state may be not finished or not finished the T task. It has: p(C = not_finished) + p(C = finished) = 1. Therefore, we have:

The T_1, T_2, . . . , T_{n-1} tasks are the prerequisite tasks of T_n task. Meanwhile the value of quantitative measure finishing level of learner to the T_n is determined by the following formula:

$$P(T_n | T_{n-1}, ..., T_2, T_1) = P(T_n | Pa(T_n)) \text{ with } Pa(T_n) \subseteq \{ T_{n-1}, ..., T_2, T_1 \} \quad (2)$$

IV. ADAPTATION ENGINE

This section represents a new approach to establish appropriate mechanisms to adapt course content to meet learner's demands. It introduces how to select the course content based on learner knowledge. In order to do that, we propose two stages: i) Quantitative knowledge level of learner to the concept as well as evaluate finishing level of learner; ii) Selection of concepts must be learned, the task should be done based on rules.

A. Select the course content based on learner's knowledge

The goal of this adaptation is to select appropriate concepts and tasks for each learner. In learning process, if learners do not complete the task, they will be guided by steps or by the component tasks that have to be taken in order to complete the task. Assume that in order to complete the course content, student needs to finish several tasks T_1, T_2, ..., T_n, and acquires the concepts C_1, C_2, . . . , C_n. If learner can finish the T_i task with his knowledge, he does not need to do the sub tasks of T_i that the system supports as guided tasks to complete the T_i task. In case of T_i unfinished, the system will guide the learner to perform some task components T_{i1}, T_{i2}, . . . ,T_{im} to complete the task T_i, the number of task components which are needed to perform, depends on different learners. Learner will be instructed to have to do the T_{i1}, T_{i2}, . . . , T_{im} and only if the T_i task is unfinished.

1) Evaluate learner's knowledge level

To evaluate learner's knowledge level, firstly we construct a BN based on course content. A set of network variables is a set of concepts, tasks, and the edges showing relationships between concept and task. We constructed a full BN as the structure of pre-defined network, with each variable in the network having a probability distribution tables. The Noisy-OR method [29] was used to construct the probability distribution table. The complexity of the method is O(k) instead of O(2^k). After that, we carry out reasoning to quantify the level of understanding concepts of learner. The goal of this step is to quantify the knowledge level of learners for each concept in each stage of learning the course, as a basis for adaptive content selection to suit each learner. We use two strategies of quantitative reasoning:

- *Diagnostic reasoning:* Going from results to causes, the evidence variables are descendants of the variables asked, denoted Q → E, where E is the

evidence variable, the variable Q is the question. This mechanism is used in cases learners do not understand a C concept, to determine the value of the probability of the learner's understanding prerequisite concepts of C concept.

- *Predictive reasoning:* Taking the results from the cause, the evidence variable is a precursor to ask variables, denoted E → Q, where E is the evidence variable, the variable Q is the question. This mechanism is used in cases of determining the quantitative value of the probability of understanding level of the C concept when the quantitative value of the probability of understanding level of prerequisite concepts is known.

The process of quantitative level of understanding learners' knowledge is done in stages during the course of study. We update the probability values of the variables in the network after the interaction with the system such as after answering a test questions. We use this mechanism by the following reasons: i) Knowledge level of learners always changes during the time they participate in the course. At each stage, the learner can only learn a part of the course content; ii) Do not update all the variables in the network, because part of course content does not cover all the concepts. This raises the efficiency of computing; iii) Quantifying the level of the understanding concept of student learning after the test aims to select the concept that learners need to learn.

2) Select concepts, tasks for each learner based on rules

Rules are basis to select concepts that learners need to learn. Therefore, we represent the adapted rules [30] through logical predicates [31]. In this step, we select concepts guiding the learners to learn as well as to point out the concepts that can be ignored. Learners are allowed to ignore concepts to learn if they have already understood the concepts. The quantitative value of the probability level of understanding of the learners, which has been identified in the previous step, is basis to determine the learners who understand the concept. The problem with the probability value is how much the learners are deemed to have understood the concept. The study by Millan [6], Wei [27], which considered the learner understanding of the concept of the probability values from 0.7 to 1, does not understand the concept when the probability value from 0 to 0.3, and was unspecified when the value of about 0.3 to 0.7. In our opinion, the choice of threshold in this model is not good because with the identification of such threshold, the concepts are equal. However, the concepts have different levels of difficulty. Therefore, the assessment of learners' understanding needs to consider the level of concepts. In our model, we determined these values based on the difficulty of the concept as in Table I.

TABLE I. PROBABILITY THRESHOLD VALUE DEPENDS ON THE DIFFICULTY OF THE CONCEPT

No	The difficulty of the concept	P(C)
1	Very Easy	[0.90 .. 1]
2	Easy	[0.85 .. 1]
3	Average	[0.75 .. 1]
4	Hard	[0.70 .. 1]
5	Very Hard	[0.65 .. 1]

V. ADAPTIVE LEARNING SYSTEM FRAMEWORK

This section is devoted to our Adaptive Course Generation System (ACGS-II). An overview of system architecture is described in the first subsection. How to apply a new approach in learner modeling, domain model and adaptation engine in this framework will be presented in the other subsections.

A. Adaptive Course Generation System Architecture

This version of our model is named ACGS-II based on ACGS [30] including three modules: Learner Module, Visualization Module and Adaptation Module as depicted in Figure 1. Learner Module is designed in order to manage learner model. Besides, it performs the evaluation, the initial classification of learners through the questionnaire, multiple-choice questions. Visualization Module manages how to represent course content to provide web interface for users. It uses several techniques to build adaptive learning such as icons, hidden links that point out the course content, which the learner can omit. Adaptation Module selects the course content in order to meet each learner's demand based on learner model.

In order to generate the course content adapt for each learners based on their knowledge. We improve three basis components of our previous framework. Specially, i) *In learner modeling*: We use overlay model to quantify the probability level of knowledge of learners. BN used to quantify the level of knowledge of learners to the concepts and tasks are considered in relation to the interdependence instead of considering the concepts and tasks independently. ii) *In content modeling*: We propose modeling the course content includes concepts and tasks. The task is underlying to the adaptive system provides instructions for each learner how to complete the task. In addition, the additional tasks to resolve problem that the course content is not only provides the pure concept, but also require learners to apply them to complete the exercises. iii) *In adaptive engine:* Our model not only to making the concept that learners must learn, but also hints the steps how to complete a task in case of the learner has not completed this by notifying direct sub tasks which learners need to do to accomplish this task.

B. Main Functions of ACGS-II

1) Create course content

This function allows teachers or course designers to declare the course content including concepts, tasks, relationships between concepts, tasks as well as relationships between concepts and tasks.

2) Create Questionnaires and Questions to test learner knowledge of course domain

This function allows course designers to declare the questionnaires, which related to course content for the initial learner classification. The questions in this section obtain preliminary information of the needs of learners, as well as some knowledge as basic of course domain. In addition, the function also allows the course designers to declare the types of test questions in order to test learners' knowledge, which is directly related to course content. The questions are the basis for assessing the learner's understanding of the subject before they participate in the course.
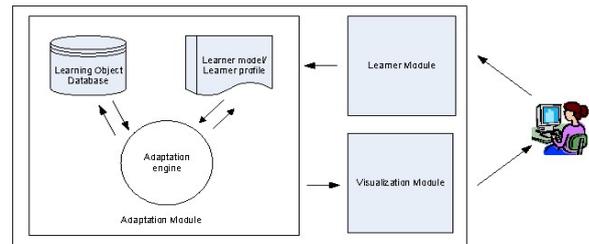
Figure 1. Architecture of Adaptive Course Generation System
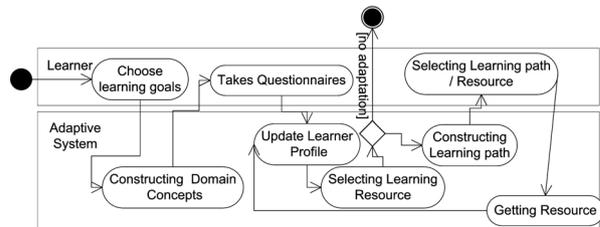
Figure 2. Adaptive Engine Operation

3) Create Exercises to evaluate the finishing level of the course

This function allows course designers to create exercises related to course content. Assignments provide assesses to the finishing level of the course. Through these exercises, the system will provide steps to guide the learner how to complete the exercises.

4) Provide appropriate learning content to each learner

This function will display the course content in accordance with the learner through adaptive mechanism.

C. Operations of ACGS-II

Operations of ACGS-II is described in Figure 2. It includes several basic steps as follows.

1) Choose Learning Goals

Learners are required to talk about their goals when they are participating in the course. The goals of the learner can be completed during the entire course, or can be learned by some concepts of the course. In this step, the learners also offer their needs, for example, time for study completion, the degree of difficulty of content, level of education to understand, analyze, or synthesize, as well as concepts, tasks to find out, etc…

2) Constructing Course Domain

Based on the learning objectives of learner, system determines the scope of the course content. It also models the course domain as a graph of concepts and tasks.

3) Take Questionnaires

In this phase, the system uses a questionnaire, the multiple-choice questions, and exercises test to survey and determine information about learners such as identifying the needs of learners, level of knowledge of the course domain. The information about learners is a basis for selecting appropriate learning content.

4) Update Learner Profile

Information on each learner system is stored and updated. The process regularly takes place during learner's participation in the course and when the learner interacts with the system such as do the test, perform the exercises.

5) Selecting Learning Resources

Based on the quantitative level of learner's understanding for each concepts and tasks, the mechanisms adapt the knowledge, and propose recommendations of concepts and tasks that learner need to learn.

VI. EXPERIMENT AND ANALYSIS

To evaluate our approach, an experiment was conducted on computer science course of a university. In the following subsections, the design and analysis of the results of the experiment are given in details.

A. Learning objectives and scenarios

The course entitle "How to design entity relationship?" was chosen by following reasons: i) there is not a course content which was built as standard data for testing. Currently, other studies selected other course domains to illustrate their research. ii) The database is a compulsory subject in information technology therefore many of learners have to participate in. Therefore, facilitate the building illustrate the learning task (one of the objectives of our study) compared with the course programming languages (C / C + +, Java...) which were chosen as illustrative examples of some systems [27], [32] due to the need of addressing each problem, requiring different solutions. To design a database, first of all, a student need to skim the problem specification and then participate in four phrases: designing entity relationship diagram, transforming entities relationship diagram to physical tables, normalizing tables, and defining queries to retrieve information.

In the model, the course content is not restricted. Course content depends on the objectives of the course, and the point of view of the course designer. Course content, however, is only used in experiment in order to illustrate our research, but not to complete content of the course. Based on the content introduced in the "Modern Systems Analysis and Design" [33], we model the course content for illustrations including 34 tasks and 24 concepts.

Students participate in the course entitle "How to design entity relationship?" via web interface. They want to skip some items of course content that they have already known rather than to learn the entire course content. How to meet student demand while they have different levels of the course domain?. System needs to point out some concepts and tasks that students can omit by providing the interface in which items were not recommended for learner is dimmed.

B. Participants

There are 500 students of information technology to illustrate and justify the research problem. It takes fifteen hours for students to finish all questionnaires, concepts and tasks.

C. Procedures

We test and evaluate ACGS-II system according to the following steps:

- Building Bayesian Network based on relationships between concepts and tasks.

- Assessing the learners' knowledge about the subject before performing the task by answering the test questions.
- Assessing the learner's knowledge in the learning process through carrying out tasks.
- Using the appropriate adaptation mechanism, given the tasks, the concepts that learner can omit based on evaluation of learner's knowledge about domain content.
- Analyzing and comparing differences in concepts and tasks that each learner need to learn in order to assess the accuracy of the test model. Comparing the results of the quantitative knowledge of the models.

1) Course domain model

The relationship between concepts and tasks of the course domain is expressed through the prerequisite relationship between concepts, tasks and relational dependencies between concepts and tasks. Figure 3 describes the requisite relationship among concepts. Figure 4 illustrate excerpt of the relationship among the tasks.
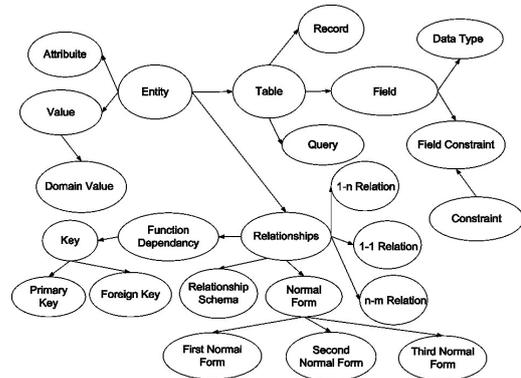
Figure 3. Excerpt of concepts relationship
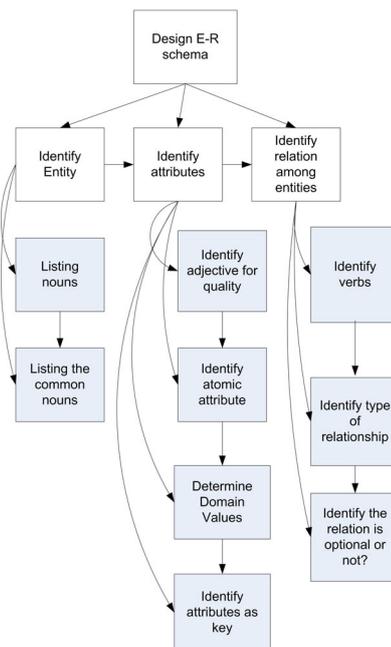
Figure 4. Except of tasks relationship

2)   Building Bayesian Network based on relationships between concepts and tasks

We built a Bayesian network, which consists of several nodes as depicted in Figure 5.

The relationship among the nodes describes the prerequisite relationships among concepts, tasks, as well as between concepts and tasks. The probability distribution table of each node is based on the experience of teachers with noisy-OR model. For instance, we illustrate the value of the probability distribution table for the node Entity concept (CE), Determination Entity (DE), listing nouns (DN), identify common noun (DCN) in Table II.

Look at Table II, it has p(DE=finished) = 0.916 when p(CE= acquired) and p(DN= finished) and p(DCN=finished).

3)   Assessing the learner's knowledge by test questions

To evaluate the preliminary knowledge of learner about course domain, the system provides some forms of multiple-choice questions to test knowledge of learners. Through these multiple-choice questions, students will be classified in the various preliminary levels. It is the basis for the adaptive learning content at the beginning. The questions are used to test the learner's understanding of the concept. Questions do not cover the entire concept graph; they only check some prerequisite concepts.

4)   Selecting concepts and tasks for each learner

The value of quantitative levels of knowledge and completing tasks is recalculated based on the dependencies among concepts, tasks through the network probability model. Based on the adaptive rules, the concepts or tasks that learners cannot learn, will be dimmed.
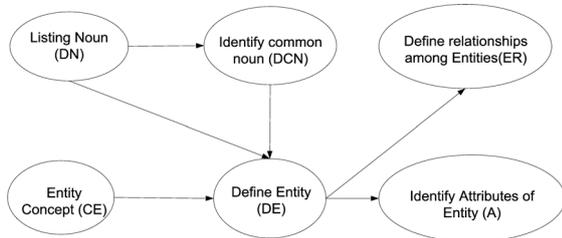
Figure 5. Bayesian network of the concept and task relation

TABLE II.   THE PROBABILITY DISTRIBUTION TABLE FOR NODE ENTITLE DETERMINATION ENTITY

CE	DN	DCN	DE	
Not_Acquired	Not_finished	Not_finished	0.0	1.0
Acquired	Not_finished	Not_finished	0.3	0.7
Not_Acquired	Finished	Finished	0.6	0.4
Not_Acquired	Not_finished	Finished	0.7	0.3
Acquired	Finished	Not_finished	0.72	0.7*0.4 =0.28
Acquired	Not_finished	Finished	0.79	0.7*0.3 =0.21
Not_Acquired	Finished	Finished	0.88	0.4*0.3 =0.12
Acquired	Finished	Finished	0.916	0.7*0.4 *0.3=0 .084

D.   Results and Analysis

The value of the initial amount of knowledge and performance of learning tasks of learners are stored in data file. For each learner, the results of assessing knowledge level of the course domain through the answer to the question are stored in files Quser01.txt, Quser02.txt,..., Quser500.txt respectively. In each file, the value measuring the knowledge level of the concepts is stored in format *conceptid: conceptid value*, in which *conceptid* is the identity of the concept of learning content model, the *conceptid value* in the range [0. . .100] means the probability value learner acquire *conceptid*. In case, the learner does not answer some questions so that the system does not assess the level of understanding of certain concepts, the *conceptid value* is denoted as *.

For example, the results of evaluate for user02 is: 18:100%; 10:*; 21:50%; 20:*; 23:*; 12:*; 14:*; 15:0% 16:100%; 34:0%; 39:50%; 40:*; 22:0%; 25:50%; 26:*; 27:*; 28:* denotes:

- User02 understood concept 18 (Entity concept) and concept 6 (Key Concept foreign)
- Probability to understand these concepts 21,15,39, and 50 of this user is 50%
- User02 does not understand these concepts 15,34, and 22
- The concept 10,20,23,12,14,26,27, and 28 not been evaluated by the User02 did not answer some questions related to these concepts.

Results of quantification of completed tasks as well as understanding concepts of learners participating in the course are stored in the file Quser01.txt, Quser02.txt,…, user500.txt respectively. In each file, the value quantifying the probability of completing the tasks is stored as format *taskid: value*. In which, *taskid* is the task identifier in the content model. The *value* in the range [0. . .100] means the probability of degree completion of the task. For example, the result of User02 is:

3:75.0

1:10.0

18:20.0

2:50.0

…

The first line shows the results evaluating when User02 performances task 3 (Identify the entity) is 75%; the next lines are results of measuring the component tasks: 1 (List of nouns), 2 (Identification of common nouns) and 18 (Learn the concept of the entity) have result 10%, 20% and 50% respectively.

Based on BN model, the value to measure p(acquired) as well as p(finish) of each concept will be recalculate. For instance, if p(*Identify the entity=finished*) = 0.75. The p(acquired) of prerequisite concepts and p(finished) of component tasks of *Identify the entity* task will be recalculated.

1: 89.656624

2: 91.02828

18: 87.185616

3: 91.8585

...

Based on the results of the experiment, we analyze statistics on several criteria:

- Dependence between the learner's knowledge of learning content before joining the course and quantity of knowledge.
- Dependence between the results of finishing the task with the amount of knowledge the learner needs to learn.
- Survey the variation probability value of completed tasks and concepts.

The first and second criteria are used to statistic the number of concepts and tasks needed to learn depending on the knowledge of learner. The third criteria examines the dependence value of the degree of the task fulfillment with the task of composition, and see the need of using probabilistic models for quantifying.

In Figure 6, the chart statistics dependency between the number of concepts and tasks, depicted as red blocks, which learner needs to learn and number of those depicted as grey blocks which can skip based on evaluating of learner's understanding of content by answering questions.

Understanding the learning content is assessed by percentage, which means that if the level of understanding is 0% so the learner does not have knowledge of the course content before joining, or the learner does not answer any test questions. Similarly, level of understanding 40% means the learner can understand 40% of the concepts that the system uses to test. In the chart, at level of understanding 100%, the system still requires learner to learn or perform 18 concepts and tasks due to insufficient number of questions in order to assess the full knowledge of the course domain and the measuring complete tasks must be assessed through actual performance.

The statistics denote number of concepts and tasks that learners are allowed to skip proportionally with the level of understanding of learners, consistent with this model system.

We statistic the number of concepts, and the tasks that learners need to learn through the implementation of the tasks. In each task, we conduct a review of dependency levels of probability of 25%, 50%, 75% and 100% completion of each task. The task includes: 3 (Identify the entity), 7 (Define attributes of entity), 13 (Identify attribute as key), 6 (Determine the relationship between entities), 24 (Define the tables), 29 (Determine the constraint), 33 (Change to First Normal Form), and 37 (Change to Second Normal Form).

In Figure 7, the chart describes the statistical results of the dependence between the number of concepts, the tasks needed to learn and level of finishing of the task. Look through the chart, if the probability measuring the finishing level of the task 3 (Identify the entity) is under 75%, so the learner will have to learn and perform 38-40 concepts and tasks. Specifically, the learners will have to perform the following tasks: Listing nouns, Listing common nouns, Identify and understand the entity concept. Otherwise, if the probability is above 75% or more, the students do not need to learn the above described tasks and concepts. Results of the statistical conclusion: the number of concepts and the tasks that learners need to learn, is inversely proportional to the degree of fulfillment of the tasks.

In Figure 8, the chart surveys the variation of the variable degree of probability evidence. With task 3, Identify the variables to be asked including task id 1(Listing the nouns), 2 (Determine common noun) and 18 (The concept of entity) which are component tasks of the task 3. Similarly, the chart in Figure 9 surveys the variation of the variable degree of probability evidence Task 6 with these tasks 22, 2 and 5.
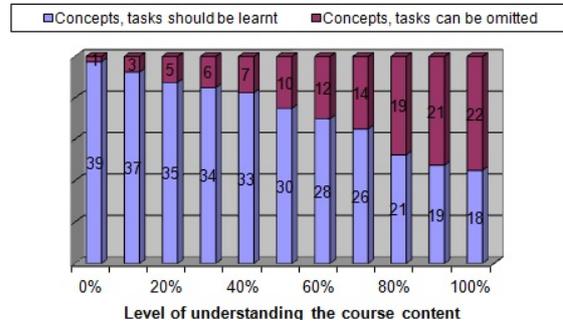
Figure 6. Task, concepts that learner can omit based on evaluation
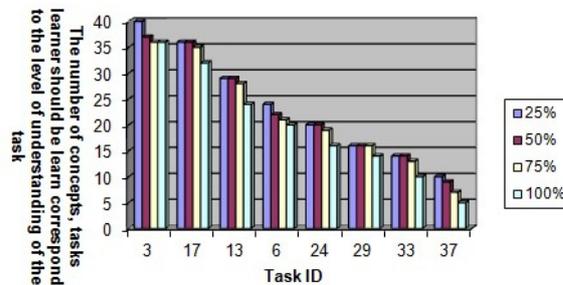
Figure 7. The dependence between the numbers of concepts, tasks needed to learn and level of finishing of the task
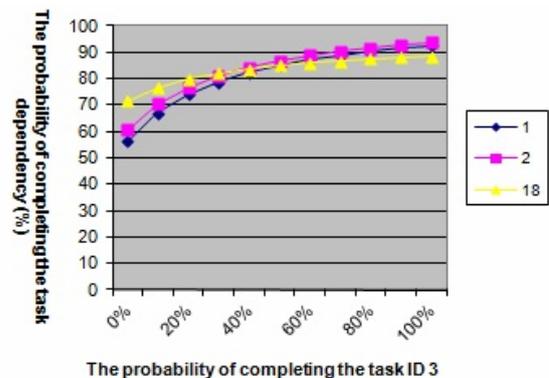
Figure 8. Variation of the variable degree of probability evidence of task id 3 with relationship tasks (1, 2, 18)
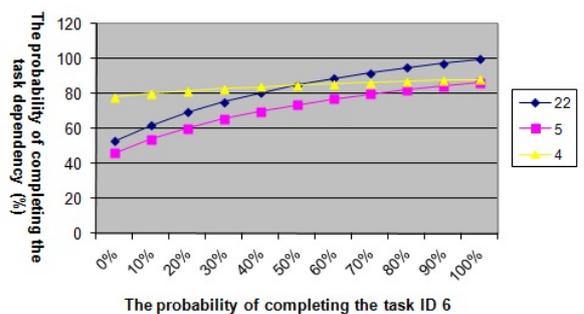
Figure 9. Variation of the variable degree of probability evidence of task id 6 with relationship tasks (22, 5, 4)

Results of the survey finding to quantify the probability of completing a task are closely dependent on the task, its prerequisite concepts. This is the basis to allow the learner whether skip the task, concept or not. The usage of probabilistic networks to quantify gives higher accuracy when using the value of qualitative or discrete.

VII. DISCUSSIONS

We represented the course domain as a set of concepts and tasks. One of the reasons of representing the course content through the task is to take into account the limitations of the current content models. These models did not focus on aspects of "How to resolve the problem?". In other words, the models did not guide learner to learn how to understand a concept, or the steps to complete a task. These models only gave the proposal to learner whether he/she should understand the concept or not. For example, to understand the C concept, learner needs to understand several concepts C_1, C_2,. . . , C_n. Based on the assessment of understanding level of C_i concepts, the system gives a value to measure level of understanding of C concept. But these models did not consider how the learner learns to understand the concept C_i.

In represented learner model, binary value such as to know or not to know how quantify the level of understanding of learners for each concept was used in adaptive learning systems [1,19] . These models do not quantify different levels of learner's knowledge of the concept. The overlay model is built in order to quantify the multiple levels of learners' understanding of concepts. Other forms of weights were used in the system including the value of qualitative, quantitative value, and probability value. Weights were used to calculate value of the discrete values as good, average, poor [15,24] which shows learner's understanding of the concept. This model facilitates the adaptation based on the rules, as well as updates the learner model. But due to limitations of the set of discrete values, it cannot classify multiple objects. In addition, the usage of discrete values has a trouble in quantifying interdependent concepts and tasks.

New approach in our model not only assesses learner's understanding of concepts such as Millan's approach [6] and Wei's approach [27] do, but also assesses the degree of completing tasks of the learner. Based on the mechanism of diagnostic reasoning and prediction of the Bayesian network, our model evaluates the completed tasks as a basis for developing the steps, which the learner should do to complete the task.

VIII. CONCLUSION

This study introduces ACGS-II - an adaptive generation system to adapt the course content based on learner knowledge by representing a new approach to learner modeling. Overlay model and Bayesian network were used to statistic and evaluate learner's knowledge level. We also constructed formulas to measure acquired concept level, the complete task level. In addition, we developed mechanisms to select adaptive course content based on evaluating quantitative values of the probability level of understanding concepts, and completing the task. Experiment results of ACGS-II were presented to illustrate the potential implementation of our environment.

Although the proposed ACGS-II provides benefits in terms of adaptive educational hypermedia system, there are several issues, which are valuable to be further research. First of all, modeling the course content as concepts and tasks as well as the relationship of them is take many teacher efforts. Secondly, quantitative level of understanding of the concepts learned in the process of taking a course as a basis to adapt the course content. Initially, the quantitative knowledge of the learner is determined through the test questions and results of exercises and tasks. In this study, we have not studied in depth the development of questions, exercises to evaluate learners' knowledge. Additionally, for students who the first time participated in, after evaluation, rather than quantitative level of understanding for each concept, the system aims to find progress in the group of users who have previously participated in the training results assessment equivalent, to provide for the beginner. Finally, developing assessment and classification model to enhance the selection of learning content for each learner is also our future research issue.

REFERENCES

[1] P.D. Bra and J.P. Ruiter, "AHA! Adaptive Hypermedia for All," *Proceedings of World Conference of the WWW and Internet. AACE*, 2001, pp. 262-268.

[2] P. Brusilovsky, "KnowledgeTree: A Distributed Architecture for Adaptive E-Learning," *Proceedings of The 13th International World Wide Web Conference*, 2004.

[3] N. Henze and W. Nejdl, "Adaptation in Open Corpus Hypermedia," *International Journal of Artificial Intelligence in Education*, vol. 12 (4), 2001, pp. 325-350.

[4] P. Brusilovsky and E. Mill n, "User Models for Adaptive Hypermedia and Adaptive Educational Systems," The Adaptive Web, LNCS 4321 Springer-Verlag Berlin Heidelberg, 2007, pp. 3-53.

[5] M. Trella, C.Carmona, and R. Conejo, "MEDEA: an Open Service-Based Learning Platform for Developing Intelligent Educational Systems for the Web," *Proceedings of Workshop on Adaptive Systems for Web-based Education at 12th International Conference on Artificial Intelligence in Education*, 2005.

[6] E. MILLALN and J.L. PELREZ-DE-LA-CRUZ, "A Bayesian Diagnostic Algorithm for Student Modeling and Its Evaluation," *User Modeling and User-Adapted Interaction,* vol. 12, 2002, pp. 281-330. http://dx.doi.org/10.1023/A:1015027822614

[7] A. Mitrovic, "An Intellignet SQL Tutor on the Web," *International Journal of Artificial Intelligence in Education*, vol. 13, 2003, pp. 173-197.

[8] N.V. Anh, N.V. Ha, and H.S. Dam, "Contructing a Bayesian Belief Network to Generate Learning Path in Adaptive Hypermedia System," *Journal of Computer Science and Cybermetics*, vol. 24 (1), 2008, pp. 12-19.

[9] C. Hockemeyer, T. Held, and D.Albert, "RATH A Relational Adaptive Tutoring Hypertext WWW-Environment Based on Knowledge Space Theory," *Proceedings of CALISCE'98, 4th International conference on Computer Aided Learning and Instruction in Science and Engineering*, 1998, pp. 417-423.

[10] N. Henze and W. Nejdl, "Adaptivity in The KBS Hyperbook System," *2nd Workshop on Adaptive Systems and User Modeling on the WWW (Toronto, Canada, May 1999)*, 1999.

[11] P.D. Bra, A. Aerts, D. Smits, and N. Stash, "AHA! Version 2.0: More Adaptation Flexibility for Authors," *Proceedings of World Conference on E-Learning*, 2002.

[12] P. Brusilovsky, "KnowledgeTree: A Distributed Architecture for Adaptive E-Learning," *Proceedings of The 13th International World Wide Web Conference*, 2004.

[13] P. Brusilovsky, E. Schwarz, and G. Weber, "ELM-ART: An Intelligent Tutoring System on the World Wide Web," *Proceedings of the Third International Conference, ITS*, 1996.

[14] N. Henze, W. Nejdl, and M. Wolpers., "Modeling Constructivist Teaching Functionality and Structure in the KBS Hyperbook System," *Proceedings of AIED99 Workshop on Ontologies for Intelligent Educational Systems*, 1999.

[15] K.A. Papanikolaou, M. Grigoriadou, H. Kornilakis, and G.D. Magoulas, "Personalising the Interaction in a Web-based Educational Hypermedia System: the Case of INSPIRE.," *User Modeling and User Adapted Interaction*, vol. 13, 3, 2003, pp. 213-267. http://dx.doi.org/10.1023/A:1024746731130

[16] G. Weber, H. Kuhl, and S. Weibelzahl, "Developing adaptive internet based courses with the authoring system NetCoach," *Third workshop on Adaptive Hypertext and Hypermedia*, 2001.

[17] J. Vassileva, "DCG + GTE: Dynamic Courseware Generation with Teaching Expertise," *Instructional Science*, vol. 26, 1998, pp. 317-332. http://dx.doi.org/10.1023/A:1003015532585

[18] P.D. Bra and J.P. Ruiter, "AHA! Adaptive Hypermedia for All," *Proceedings of World Conference of the WWW and Internet. AACE*, 2001, pp. 262-268.

[19] P. Brusilovsky, J. Eklund, and E. Schwarz, "Web-Based Education for All: A tool for Developing Adaptive Courseware," *Proceedings of Seventh International World Wide Web Conference, H.A. Thistewaite*, 1998, pp. 291-300.

[20] M. Specht and R. Klemke, "ALE - Adaptive Learning Environment," *Proceedings of WebNet'2001, World Conference of the WWW and Internet. AACE*, 2001, pp. 1155-1160.

[21] J. Willamowski, F. Chevenet, and J.M. François, "A development shell for cooperative problem-solving environments," *Mathematics and computers in simulation*, vol. 36, 1994, pp. 361-379. http://dx.doi.org/10.1016/0378-4754(94)90070-1

[22] J. Oberlander, M.O'Donell, C.Mellish, and A.Knott, "Conversation in the Museum: Experiments in Dynamic Hypermedia with the Intelligent Labeling Explorer," *The New Review of Multimedia and Hypermedia*, vol. 4, 1998, pp. 11-32. http://dx.doi.org/10.1080/13614569808914693

[23] S. Ritter, "PAT Online: A Model-Tracing Tutor on the World-wide Web," *Proceedings of 8th World Conference on Artificial Intelligence in Education*, 1997, pp. 11-17.

[24] C. Choquet, F. Danna, P. Tchounikine, and F. Trichet, "Modeling the Knowledge-Based Components of a Learning Environment within the Task/Method Paradigm," *Proceedings Intellegent*

Tutoring System Conference, 1998, pp. 56-65. http://dx.doi.org/10.1007/3-540-68716-5_11

[25] N.V. Anh and S.D. Ho, "ACGs: Adaptive Course Generation System - an Efficient Approach to Build E-learning Course," *Proceedings of the IEEE Sixth International Conference on Computers and Information Technology*, 2006, pp. 259-265.

[26] N.V. Anh and S.D. Ho, "Applying Weighted Learning Object to Build Adaptive Course in E-learning," *Learning by Effective Utilization of Technologies: Facilitating Intercultural Understanding, Frontiers in Artificial Intelligence and Applications*, 2006, pp. 647-648.

[27] F. Wei and G. D.Blank, "Student Modeling with Atomic Bayesian Networks," *Proceedings of International Conference in Intellegent Tutoring System, LNCS 4053*, 2006, pp. 491-502.

[28] C. Zhao and L. Wan, "A Shortest Learning Path Selection Algorithm in E-learning," *Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies (ICALT 2006)*, 2006. http://dx.doi.org/10.1109/ICALT.2006.1652375

[29] J. Pearl, Probabilistic Reasoning in Expert Systems: *Networks of Plausible Inference, San Francisco*: Morgan Kaufmann Publishers, Inc, 1988.

[30] N.V. Anh, N.V. Ha, and H.S. Dam, "Developing Adaptive Hypermedia System Based on Learning Design Level B with Rules for Adaptive Learning Activities," *Journal of Natural Science, Vietnam Nation University*, vol. 25(1), 2009, pp. 1-12.

[31] M. Kiefer, G. Lausen, and J. Wu, "Logical Foundations of Object-Oriented and Frame-Based Languague," *Journal of ACM*, vol. 42, 1995, pp. 741-843. http://dx.doi.org/10.1145/210332.210335

[32] N. Henze, "Adaptive Hyperbooks: Adaptation for Project-Based Learning Resources," 2000.

[33] J.A. Hoffer, J.F. George, and J.S. Valacich, "Modern Systems Analysis and Design," Addison Wesley, 1998, pp. 641-682.

AUTHORS

Viet Anh NGUYEN is with the Center of Computer Network and E-learning, University of Engineering and Technology, VNU, E3, 144 XuanThuy, CauGiay, Hanoi, Vietnam (e-mail: vietanh@vnu.edu.vn).