Chapter 6

Conclusion and Future Work

The application of decision theories and Bayesian network concepts to Intelligent Tutoring System design, for the purpose of modeling the student and implementing pedagogical theories has been studied in detail in this thesis. While having acknowledged limitations in some cases, the approach followed is a powerful new approach to Intelligent Tutoring System architecture design. Section 6.1 outlines some of the original contributions of this thesis towards the computerized tutoring. Future research directions are outlined in Section 6.2. Final remarks are given in Section 6.3.

6.1 Summary of Contributions

The main problem associated with the current Intelligent Tutoring System is uncertainty in evaluating the student's knowledge. This thesis contributes towards minimizing uncertainties prevailing in various forms in Intelligent Tutoring System. Powerful general theories of decision-making, designed specifically for situations involving uncertainty, have been developed. The other problem associated with Intelligent Tutoring System, namely, the ineffective use of test items, can be addressed by the usage of item-response theory. This research focused on representation of the knowledge structure as Bayesian networks, decision-theoretic approach to computerized tutoring, the application of item-response theory for item selection and the generation of tutoring policy. As a result of this research, an intelligent decision-theoretic tutoring system, BiTutor, has been developed. The implementation of BiTutor and its evaluation (based on a group of students) show that
decision-theoretic formulation of tutoring can be applied to various knowledge domains. The following subsections summarize the four key contributions of this research.

6.1.1 Representing Knowledge Structure as Bayesian Networks

In this research, the knowledge component to be represented is in the form of topics, subtopics, and concepts (Knowledge Items) that are written in the course syllabus. Each knowledge item represents what is to be taught to the student and at which level of the cognitive domain (McCormick & Pressley, 1997). Each knowledge unit can be represented as a node in Bayesian network. The random variables in Bayesian networks are used to model uncertainty of student's knowledge mastery of the knowledge node in the curriculum.

The relationship among variables in the knowledge structure can be represented using the conditional dependencies of the variables in the Bayesian network (Pearl, 1988). In this research, the approach of using the knowledge items and their causal relationships to develop a set of Bayesian networks has been adopted. Such network is an efficient representation of hierarchical information, where inclusion of new information only affects local conditional relationships. Bayesian networks provide easier mechanism for updating the knowledge structure. This research uses both the pre-requisite and belong-to relationship in a single layer. One of the main contributions of this work is to construct the conditional probability tables using the weighted approach.

6.1.2 Decision-theoretic Approach to Computerized Tutoring

Probability theory describes what an agent should believe on the basis of evidence; utility theory describes what an agent wants, and decision theory puts the two together to describe what an agent should do (Russell & Norvig, 2003). Another contribution of this work is to apply the decision theory in Intelligent Tutoring System for action selection. Decision

Conclusion and Future Work
analysis provides a philosophy that emphasizes the insights that can be gained by decision maker who goes through the decision analytical processes.

The aim of the decision-theoretic approach in a computerized tutoring system is to provide optimal action selection which maximizes student learning and is defensible. Pedagogy can be incorporated into decision analysis to measure learning value, besides taking into consideration uncertainty in student's knowledge mastery. In this way, one can be sure that the correct standard for tutoring decisions can be achieved. After an action is decided, other consequential actions, such as, which topic to be presented next and which item to be used can be determined. The utility functions are formulated according to the student's behavior for each tutoring action. A systematic framework for making tutoring decision under uncertainty, in which the feasible alternatives are identified, has been formulated in this thesis.

6.1.3 Application of Item-Response Theory for Item Selection

Assessment has always been a very important step in the learning process. In most of the Intelligent Tutoring System, assessment is done by posting items to the students and tracking their responses to these items. These are usually in multiple-choice format and the questions can be designed such that each item tests one knowledge item. To make the system adaptive, it is necessary to rank the items according to their difficulty and also the statistical measurement of student's mastery level is needed. The third major contribution of this thesis is the formalization of an innovative method for selecting challenging items. Item selection has been done by using the concepts of item-response theory where each item is defined using a set of parameters. These item parameters can be estimated based on student's responses using item calibration methods such as maximum likelihood method, Markov Chain Monte Carlo (Carlin & Chib, 1995; Gilks et al., 1996), Kernel Smoothing.
(Ramsay, 1991) or Method of least squares. This results in the ability to form probabilistic relations between the key concept mastery states and item difficulty levels.

Each item is associated with certain feedback. If a student gives a wrong answer, corresponding feedback should be provided to the student. With these features, the student model can be constantly updated with student's mastery level and feedback on his misconceptions.

### 6.1.4 Generation of Tutoring Strategy

The fourth major contribution of this research is that optimal tutoring strategy can be generated in a polynomial time. Student's learning is seldom completed with one tutoring action. A tutoring strategy or policy consists of a planned sequence of actions to guide student learning. The system selects the best tutoring action based on student's current mastery state. The response affects the student's next mastery state. The goal is to determine actions that seek to maximize information about the student's misconceptions or faulty knowledge. This leads to the possibility of personalized tutoring. This sequence of actions leads to the shortest possible time. These actions tell the tutor what to do for any state that the student might be in.

Before taking actions, the complete information on the variables in the student model is not available and so the decision-theoretic tutoring is a Partially Observable Markov Decision Problem (PODMP). In PODMP (Lane, 1989; Cassandra et al., 1994; Russell & Norvig, 2003), decisions are made by projecting forward a sequence of possible actions and choosing the best one. For evaluating the effectiveness of BiTutor and their satisfaction with the pedagogy, data obtained from a group of students are used. This has been done by conducting pre- and post-test for two groups of students. Students in the e-learning group...
have access to BiTutor and are supervised by a staff, while the students in the control group are taught in the traditional way (without BiTutor). The results indicate that BiTutor has substantially improved the students' mastery of the domain knowledge.

6.2 Future Work

BiTutor has been developed as a decision-theoretic platform which enables the future study on extensions of decision theory in computerized tutoring. Some of the worthwhile future works are described in the following subsections.

6.2.1 Refinement of Bayesian Network's Structure by Learning from Data

In BiTutor, experts construct the Bayesian network based on their knowledge on the subject. After BiTutor has been in use for a period of time, the data collected from the student's interaction with the system would be sufficient to refine the Bayesian networks in terms of conditional dependent relationships and probability values. There are a number of benefits with this data-centric approach. Firstly, because the model is induced from actual data, its predictive performance can easily be evaluated by testing the network on data that was not used to train. Secondly, data-centric models can be expected to be much smaller than the typical expert centric model because the latter represents both observed and hidden variables, while the former models only observable variables. Usually Bayesian network structures involve a fairly large number of nodes. However, inaccuracies or changes might affect only some subset of the variables in the network. Also, when the data collected about the domain is partial (Information only about a subset of network variables), it is possible to refine the structural relationships that exist between the variables mentioned in the data collected. Tutoring can be made more adaptive by following this approach.
6.2.2 Use of Three-Parameter Logistic Item Response Model

For multiple choice test items, cognitive theory suggests that when an examinee does not know the correct response, the individual will guess. In situations where guessing is possible, the assumption $\lim_{\theta \to \infty} P(\theta) = 0$ is not a reasonable assumption of the cognitive process the model is attempting to measure. For this reason Birnbaum (1968), developed a generalization of the two-parameter logistic that allows the $P(\theta)$ to have a lower asymptote different from zero. The generalization is

$$P_j(\theta) = c_j + \frac{1-c_j}{1+e^{-\alpha_j(\theta-b_j)}}$$

(6.1)

The three-parameter logistic assumes that the examinee knows the correct answer of the item with probability equal to that of two-parameter logistic function or guesses the item correctly with probability $c_j$.

Also, the student's ability used in two-parameter logistic should be changed to take into account two kinds of information: the first one includes information that does not change over the learning process, like learner's capabilities (degree, background knowledge,...), learner's technical characteristics (computer expertise, connection speed,...), learner's preferences (learning style, screen options,...), etc. This information can be collected at the beginning of the learning process using forms and tests. The second includes those that change over the learning process, like the learner's knowledge level for each knowledge unit, skills, goals, etc.

In BiTutor, for estimating the item parameters, the method of least squares is used which can not be applied for three-parameter logistic functions. For estimating the parameters of three-parameter logistics, other standard parameter estimation procedures such as marginal maximum likelihood (Bock & Aitkin, 1981), joint maximum likelihood (Baker, 1992) or...
Markov Chain Monte Carlo (MCMC) (Carlin & Chib, 1995; Gilks et al., 1996; Patz & Junker, 1997) are available.

6.2.3 Provision for Quick Tutorials

BiTutor is currently being developed using JAVA, XML and JSP programs so as to support self-paced learning in World Wide Web (WWW) environment. As it is being developed as a web-based application, the system is available for all the internet users. A student can use the system to learn any subject available in BiTutor and can expertise the same. As BiTutor identifies pre-requisite relationships in mastering a topic, pre-requisites must be learnt before learning the actual required topic. This style of learning will be suitable for students learning a topic from the beginning. But, for a student who wants to refresh his knowledge on a particular topic or wants to master a topic alone, quick tutorials must be provided. In quick tutorials, even if there are some pre-requisites, only the necessary concepts from the pre-requisites will be taught to the students. This makes the learning process faster as the students can pick the topic they want to learn.

6.3 Final Remarks

Intelligent Tutoring System technology is currently in its infancy. Systems developed in the laboratory have found their way only in recent years into the classroom, but this is usually only for short-term evaluations. Despite this, there is a large potential commercial market for quality Intelligent Tutoring Systems, especially now that large-scale low-cost electronic distribution of software is possible via the Internet. The main differentiating factor between Intelligent Tutoring Systems and other educational software is the built-in "intelligence"; in addition to exercises and multi-media tuition, the Intelligent Tutoring System has a pedagogy which it applies to make learning more efficient. Research should therefore be
directed towards making the pedagogy more effective. Indeed, the community of student modeling researchers is going a long way towards achieving this. The contribution of this thesis is to show that decision-theoretic approach is an effective tool for computationally representing and implementing pedagogical theories.