THE ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
FASCICLE III, 2007 ISSN 1221-454X
ELECTROTECHNICS, ELECTRONICS, AUTOMATIC CONTROL, INFORMATICS

KNOWLEDGE BASED DIAGNOSTIC SYSTEM WITH PROBABILISTIC APPROACH

Adina Cocu

*University "Dunarea de Jos" of Galati, Department of Computer Science,
111 Domneasca Street, Galati, Romania*

Abstract: This paper presents a knowledge learning diagnostic approach implemented in an educational system. Probabilistic inference is used here to diagnose knowledge understanding level and to reason about probable cause of learner's misconceptions. When one learner takes an assessment, the system use probabilistic reasoning and will advice the learner about the most appropriate error cause and will also provide, the conforming part of theory which treats errors related to his misconceptions.

Keywords: diagnostic system, educational system, probability theory, graphical representation

1. INTRODUCTION

Currently, many computer assisted instructional systems use artificial intelligent techniques, such as fuzzy techniques, Bayes and neural networks (Katz et al, 1992; Grigoriadou et al, 2002; Guzman, Conejo, 2004; Stathacopoulou and all, 2004) in order to assist the learner in the instructional process. Our system proposes a diagnostic mechanism based on probabilistic reasoning techniques for the uncertainty diagnosis of knowledge understanding level.

The instructional process consists in creation of the new structures in the human brain. The human nervous system is made by neuronal cells that form networks capable to transmit and to process information. Some of the networks are inherited and others are created in the instructional process. The new structures created in learning process became stable through repeatability of stimulus-response pair.

In order to achieve correct and complete knowledge, the learner must be guided in the progression of learning. The objectivity of this process imposes the usage of computer diagnoses techniques in obtaining, as good as possible, precise and adapted information about learner knowledge level.

The knowledge about learning domain and about learning persons are obtained from real world and these are subject of imperfection. Thus, the intelligent tutoring systems must be able to represent and to reason with imperfect knowledge in order to better adapt to the learner necessities.

The system proposed by us has two main features. First, it uses probabilistic inferences techniques to perform a measure of the most probable situation of learner knowledge understanding. Secondly, the system can automatically create a diagnostic and an explanation about the learning domain parts necessary to be repeated by the learner.

2. KNOWLEDGE BASED DIAGNOSTIC SYSTEM

The domain knowledge of a Language Programming course is modeled in this study. Experts' knowledge contains theoretical and practical concepts to be trained to the learner. This learning domain is structured through the different levels modules and a

This paper was recommended for publication by Severin Bumbaru

study curriculum is made by modules combination. Each module contains sets of basic and advanced learning concepts presented into relational knowledge representation structure. These concepts can be represented graphically and implemented in software made by us. Each node from the graphical structure represents a concept in the domain knowledge. The relation between concepts is represented like an edge among nodes. The relationships strength between concepts is represented through probabilistic values. The concepts from the learning domain are represented with a directed acyclic graph similar with Bayesian network. This structure can be implemented in practice and can be saved like a file with a structure of tags. Each tag represents a node or a relation between nodes or an associated attribute (number of children, number of parents, probability measures).

The system uses the probabilistic knowledge represented graphically for learning diagnosis purposes. The diagnose process regards the learner current learning state and allows users (learner or tutor) to identify the potential erroneous concepts of knowledge learner.

The diagnose process may have the following action plan: the system presents questions to the student, picks up his answers and examines the correctness of each answers. When the student's answer is erroneous, the system attempts to diagnose the underlying misconception of the mistake.

The learner makes an assessment test and the results of this test are compared with expert's solution. Then, the results are introduced as base for evidences in system. The differences between students' and experts' knowledge can be viewed like errors or misconceptions. Determination of differences is the aim of diagnose system. When the evidence is complete, the system can make probabilistic inference of student knowledge learned in the interest domain area. Figure 1 shows an example of graphical representation, modeling four variables concepts for the course: predefined data types (T), variables (V), operators (O), control structures (S).

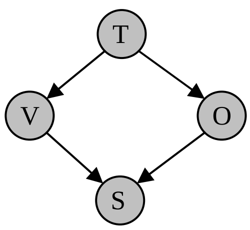

Fig. 1. An example of variables represented graphically.

The next four tables show the prior probability for independent variable T and conditional probabilities distribution for variables V, O and S dependent on combined values of parents. Each variable can have two probable values: known (K) and unknown (UK). The values of these probabilities can be obtained from expert in the learning domain or from the tutor, based on their experience. The probability measures from each line of each table must satisfy the normalization conditions, means their sums must be one.

Table 1 Probabilities for concept T (predefined data types) for the two possible values

K	**UK**
0.5	0.5

Table 2 Distribution of probabilities for concept V (variables) for the two possible values dependent from the values of parent T

T	**K**	**UK**
K	0.8	0.2
UK	0.4	0.6

Table 3 Distribution of probabilities for concept O (operators) for the two possible values dependent from the values of parent T

T	**K**	**UK**
K	0.8	0.2
UK	0.1	0.9

Table 4 Distribution of probabilities for concept S (control structures) for the four possible combination of values of parents V and O

V	**O**	**K**	**UK**
K	K	0.9	0.1
K	UK	0.2	0.8
UK	K	0.4	0.6
UK	UK	0.1	0.9

In order to test the system, in diagnostic process, evidence about variables S and O is introduced in system. These evidences are obtained from the results of an assessment session take by the learner. Suppose that the learner do not know to deal with control structures and neither with operators. Thus, the values for variables S and O are unknown (UK).

3. PROBABILISTIC APPROACH

In Bayes network theory, when there are evidences about enounces then it is used evidential reasoning. Evidence values for some variables involve modification of probability measure for corresponding values. For example, probability of variable S taking value UK is P(S=UK|O, V)=1, independently from values of parents O and V. So, in consequence, P(S=K|O, V)=0.

First step of reasoning in Bayes networks consist in calculation of marginal probabilities for each variable from graph. In this case it is used equation (1) based on distribution of probabilities for all variables. Thus, in our study, there are two marginal probabilities for variable V, because there are two probable values of V.

$$P(V=K)=\sum P(T, V=K, O, S) \quad (1)$$

Probability from the sum in equation (1) is named joint probability for the entire set of variables. In order to calculate joint probability it is used next formula (2).

$$P(x_1,...,x_n) = \prod_i P(x_i \mid pa_i) \quad (2)$$

Where $x_1,...,x_n$ are the variables from network representation and pa_i are the parents of x_i variable.
In this study, the initial marginal probabilities calculated for all variables are presented in table 5.

Table 5 Initial marginal probabilities for all variables

	T	V	O	S
K	0.5	0.45	0.6	0.46
UK	0.5	0.55	0.4	0.54

The system makes automated inference in order to propagate evidences (after the user introduces these evidences). In the situation enounced before, for S and O variables, the recalculated marginal probabilities are shown in table 6.

Table 6 Marginal probabilities for all variables recalculated after introducing new evidence

	T	V	O	S
K	0.5	0.45	0	0
UK	0.5	0.55	1	1

The diagnostic reasoning means finding the causes based on the effects. In the case discussed in this article, the effects are those that the student does not know the concepts associated with operators and control structures.

The aim of our diagnostic system is to find the causes for deficiencies in learned concepts of the learner. In our directed graph representation it is considered that parents are the causes for their children. Thus, the directed causes for S are O and V, and for the O is T. But in reality of educational systems, errors and mistakes do not have just directed causes. The learning errors can be combined between them and a wrong answer may have multiple diagnoses or there is not a unique answer for an item.

As we can see in table 6, the values for marginal probabilities for variable T and V after propagation of new evidence are the same like initial values viewed in table 5 (columns two and three from both tables). So, there is no informational gain. This fact is not very helpful for diagnostic purposes, because there are no differences between values before and after introducing the new evidences. In these conditions, our diagnose system propose another probabilistic approach.

This approach sets out from the idea that, in educational systems, errors diagnose is a complex task. The system must provide to the user the needed help for misconceptions corrections. The method has to explain to the learner what are the commonly mistakes and what are the concepts that must be repeated for raising his understanding level. In other words, what are the concepts of background needed to be learned in order to make a better score to the assessment process?

The new approach proposed in this diagnostic system is based on calculation of joint probability using equation (2). These values reflect the behavior of all variable from graphical representation. To facilitate the diagnostic purposes, the joint probability measures must be computed for all possible values of all variables. Also, it is important that this computation reflect the new evidence introduced in the system. In the above considered example, the proposed inferential process calculates the joint probability for all possible values combination. The evidence discussed on top of this chapter refers only to variables O and S. In this situation the joint probability calculus is made only for combination of values for T and V variables.

In this work, we propose a method that chooses the biggest joint probability measure to indicate the most probable state from all possible combinations. In table 7 it is shown this value and the combination of variable values for the evidences case presented above.

Table 7 The biggest joint probability for all variables computed after introducing new evidence

Joint measure	**T**	**V**	**O**	**S**
0.45	UK	UK	UK	UK

The explanation of this result is that the user must repeat concepts about predefined data types and variables when he doesn't know operators and control structures concepts, because this is the most probable situation. In our study, a human tutor asked about this situation, gives the same guidance in the same conditions.

The proposed system include an advice generator component that tries to respond to the learner in the most appropriate way about possible error cause and also provides the conforming part of theory which treats errors related to the subject. For example, "You must repeat concepts T and V find in chapter 1 and 2".

3. CONCLUSIONS

Usually, a computer assisted instructional system contains a diagnose module that has the aim to support the student in errors explanation through an advise component. Corresponding to each type of assessment item, the intelligent tutoring system contains libraries with different learning concepts and associated mechanisms for supplying the most appropriate explanation to the student's mistakes. This is the reason why this study tries to improve the performance of diagnostic module with probabilistic approach.

This study was continued with more a complex concepts graphical representation. The gathered information obtained from experiments was used to evaluate probabilistic inference against tutor classical method. In result, the output of the diagnostic module concerning student knowledge level, compared with the assessment of a tutor, gives similar results, in the most cases. In conclusion, this probabilistic approach has great opportunities to be used in diagnose knowledge understanding of a learner.

The disadvantage of this approach is the increase of the calculus complexity when increase the number of concepts.

The probability theory is capable to manage uncertainty and it is near by the human treatment of uncertainty. The reason why we choose this probabilistic approach it is the power obtained from combining clarity of graphical representational formalism with the stability of probability theory.

REFERENCES

Grigoriadou M., Kornilakis H., Papanikolaou K., Magoulas G., (2002), "Fuzzy Inference for Student Diagnosis" in *Adaptive Educational Hypermedia*", in Vlahavas I.P. and Spyropoulos C.D Eds., Springer-Verlag, Berlin.

Guzman E., Conejo R., (2004), "A Model for Student Knowledge Diagnosis Through Adaptive Testing", Springer-Verlag.

Katz S., Lesgold, A., Eggan G., Gordin M., (1992), "Modeling the student in Sherlock", *Journal of A.I. in Education*, pp. 495–518.

Stathacopoulou R., Magoulas G. D., Grigoriadou M., Samarakou M., (2004), "Neuro-fuzzy knowledge processing in Intelligent Learning Environmets for Improved Student Diagnosis", *Information Sciences*, Elsevier.