Bernardo Quibiana

Shamik Roy Chowdhury

CSCE 582 – Fall 2010

October 20, 2010

A Summary of “Bayesian Networks in Educational Testing” by Jiri Vomlel

In this paper, Vomlel discusses the diagnosis of the presence or the absence of person’s skill through Bayesian networks, and he argues that modeling the dependence between skills leads to faster and more informative diagnosis. The test design initiates by identifying a set of tested skills, abilities and misconceptions and a bank of questions , such that “only skills from set are directly related to question”. The skills and questions are represented by random variables having finite sets of values, and their dependence can be modeled with a Bayesian network as shown in *Equation 1*. The models consist of one student model, which models the relation between skills, one evidence model, which models the relation between skills and questions, and an overall model which combines the previous two models.

 (*Equation 1*)

Vomlel discusses the construction of both fixed tests (consisting of a fixed sequence of questions) and adaptive tests (consisting of a sequence of questions, where the next question is selected according to the answer to the current question). The value-function *entropy* (Equation 2) was introduced to estimate the uncertainty of the hypothesis (lower entropy implies less uncertainty), and thus outline an optimal test. To maximize the test information, Vomlel uses a myopic approximation, minimizing the expected value of the entropy with the answer to each question. Further, Vomlel designed an algorithm for constructing an optimal fixed test that uses entropy as the parameter to search for the most informative question at each stage of the search. In an adaptive test, the total expected entropy on skill is used a criterion to select the next question.

 (*Equation 2*)

To show empirical evidence, Vomlel modeled tests on the domain of fractions done by a group of students from Aalborg University. The students prepared two paper tests each with 10 groups of exercises, and 149 students from Brønderslev High School took the tests (resulting in 149 data vectors). The tests diagnosed three groups of basic fraction skills (see Table 1) and one group of misconceptions based on the results of the tests. Vomlel used the Hugin PC-algorithm to search model structures and modified the obtained models according to the expert knowledge, resulting in nine different models. Next, he used a leave-one-out cross-validation procedure (learn the models on 148 data vectors and test them on the remaining one) to test the models’ skill predictive accuracy (percentage of skills whose most-probable state equals the observed state). Furthermore, Vomlel compared four test design methods: a test with questions ordered as in the paper test given to the students, a test with questions ordered in the reverse order of the test given to the students, an optimal fixed test constructed with Vomlel’s algorithm, and an adaptive test constructed using the student model.

|  |
| --- |
| Table 1. Basic fraction skills diagnosed in the paper tests |
| Elementary skills | Operational skill | Application skills |
| Comparison | Finding common denominator | Finding common denominator |
| Addition | Cancelling out | Cancelling out |
| Subtraction | Conversion to mixed numbers | Conversion to mixed numbers |
| Multiplication | Conversion to improper numbers | Conversion to improper numbers |

Vomlel found results in favor of modeling the dependence between skills. In all test design methods, the skill predictive accuracy increased with the number of questions answered. However, the first two test design methods performed poorly, while the adaptive test had the best predictive accuracy followed by the optimal fixed test. Finally, the adaptive test is the most informative test with least number of questions.