ADAPTIVE TEST USING BAYESIAN NETWORK

Using Questions from Fractions and Numbers.

Table of Contents

1 INTRODUCTION TO THE TASK 1

2 BAYESIAN BELIEF NETWORKS 2

2.1 Use of Probability Theory 2

2.2 Bayesian Belief Network 3

2.3 Expert Centric Bayesian Student Model 4

2.4 Conditional Probability Tables in Bayesian Model 5

3 THE ADAPTIVE TEST BAYESIAN NETWORK MODEL 7

3.1 Skills Model 7

3.2 The Complete Model 8

3.3 The Question Set 9

3.4 Defining Each Question 9

4 ITEM SELECTION FOR ADAPTIVE TESTING 11

4.1 Psuedo code for entropy calculation 12

5 APPENDIX 13

5.1 Screen Shots 13

5.2 Source Code 14

5.2.1 Heuristics.cpp 14

5.2.2 Init.cpp 16

5.2.3 Wrappers.cpp 18

5.2.4 Gui.cpp 19

5.2.5 Gui.h 27

5.2.6 Global.h 28

5.2.7 Heuristics.h 29

5.2.9 Wrappers.h 29

1 INTRODUCTION TO THE TASK

One task in educational testing is diagnosis of the presence or the absence of person's skills. A possible scenario is that an institution provides a certain course and wishes to test the applicants for gaps in their prerequisites for attending the course. This task is more difficult than giving grades to students or the classification of examinees according to their overall level since the goal is not only to say whether an examinee is good or not but to pinpoint each examinee's abilities and weak points.

The Aim of the project is to implement an adaptive testing technology using Bayesian Network. This adaptive test was designed as such that it could pinpoint the student's weakness or strength in different sub-domains of the considered subjects or field.

To create the test we took question sets from Elementary Fractions taught in the Mathematics course of grade 5 and 6 students. The model for the subject (in our case Elementary Fractions) was created manually (since we college students consider ourselves expert enough to create question on this field target grade 5 students). These types of models are referred as expert centric Bayesian model. We will introduce Bayesian Models and their type in later chapters.

2 BAYESIAN BELIEF NETWORKS

2.1 Use of Probability Theory

Intelligent tutors cannot discover students ′ exact knowledge about specifi c topics. They might ask students questions about a topic and additionally ask them about related topics in order to determine their knowledge. Yet often tutors must infer knowledge about a specifi c topic from knowledge of related topics. They might model a partial ordering among topics as a way to learn about student's knowledge of specific topics (Reye, 2004). Assume that the tutor intends to discover whether a student knows about topic S and that topic R is a prerequisite for knowing topic S, such that a lack of knowledge about R implies a lack of knowledge of S. A tutor might reason about student's knowledge of topics R and S as shown in Figure 7.2 , in which variables are presented as nodes and relationships as arrows between nodes. When uncertainty is not included and only fi rst-order predicate logic was used, this prerequisite relationship was expressed as follows:

Understand_topic (student, S) → Understand_topic (student, R)

or equivalently

Not_understand_topic (student, R) → Not_understand_topic (student, S)

However, the situation is more complex than predicate logic can represent. One might express the fact that additional evidence of a student ′ s knowledge of R causes the tutor to revise its belief about student knowledge of S (Reye, 2004). This latter statement cannot be modeled in predicate logic as it involves reasoning under uncertainty. The tutor cannot know with certainty that the student knows S even when we state that the student ′ s knowledge of R provides strong evidence that the student also knows S (Reye, 2004). Belief networks provide a way to visualize our reasoning about prerequisite relationships and related beliefs, expressed as probabilities. The probability that a student knows S given that we know that he does not know R is stated as a conditional probability:

P (student_knows (S) | student_knows (R)) = 0

We also state that the probability that a student does not know S given he does not know R is:

P (student_knows (S) | student_knows (R)) = 1

To indicate a close relationship between knowledge of R and S, we write:

P (student_knows (S) | student_knows (R)) =0.95

Figure 2-1 Simple reasoning about related beliefs.

or that a student who knows R is highly likely (95%) to also know S. Alternatively, we assert that knowledge of R has little impact on the likelihood of knowledge of S and the ***prior probability*** of student knowledge of S is 0.01:

P (student_knows (S) | student_knows (R)) =0.01

Probabilistic inference does not work like logical inference. It is tempting to interpret the statement P (angles | exercise k) _ 0.80 to mean that " whenever students complete exercise k, we conclude that the probability that they know angles (to calculate the angles in a triangle) is 80%. " This is wrong on two counts. First P (Skill i) always denotes the *prior probability* of the skill being known (enumeration of belief data prior to testing the student), not the *posterior probability* (evidence or belief after observing the data). Second the statement P (angles | exercise k) _ 0.80 is only applicable when exercise m is the only available evidence. When additional evidence, such as exercise m , is also available, we must calculate P (Si | E k ∧ E m). Figure 7.3 suggests that a student's knowledge of C depends on knowledge of both A and B. Clearly the use of conditional probabilities extends to longer and denser chains of networks.

2.2 Bayesian Belief Network

Bayesian theory can roughly be boiled down to one principle: To see the future, one must look at the past (Leonhardt, 2001). Bayesian methods reason about the probability of future events, given their past and current probabilities. They are based on the understanding that the world is rife with uncertainty and often not suited to clean statistical tests . Rev. Thomas Bayes, an 18th-century cleric who lived from 1701 to 1761, offered a mathematical formula for calculating probabilities among several

Figure 2-2 Reasoning about related beliefs extended to multiple

variables that were causally related but for which the relationship was not easily derived by experimentation (Bayes, 1763). Bayesian belief networks (BBNs) enable computers to combine new data with prior beliefs about data, make subjective decisions about how strongly to weigh prior beliefs, and provide a policy for keeping new information in the proper perspective (Leonhardt, 2001). They provide a graphical method to design probabilistic models based on conditional probabilities and the Bayes formula. BBNs are enabled by improved computer power and development of key mathematical equations in the late 1980s that supported networks with enough variables for practical applications.

2.3 Expert Centric Bayesian Student Model

There are three different Types of student model that are widely used these days. The classification has been done according to whether they were built by experts, learned through data, or both. Here we will only discuss the Expert Centric Data model in details. This is the class of data model we have used in our adaptive Test.

1. In this category of student model BBN is *expert-centric* or networks and conditional probabilities are specified either directly or indirectly by experts. Experts create the network structure or topology, draw the arcs, and define the conditional probability of the arcs.

One disadvantage of expert-centric student models is that any student has hidden states (e.g., emotion) and parameters that are diffi cult to infer (Mayo and Mitrovic, 2001). Suppose the tutor maintained the conditional probability P (Problem 023 | KnowNewton'sLaw) for computing the posterior probability (evidence or belief after observing the data) distribution over a database of many students. Consider how this model might be adapted to an individual student (Mayo and

Mitrovic, 2001). There are two approaches. The fi rst is to observe student action while solving many problems, instantiate these outcomes, and then update the value of the student's knowledge.

The BBN structures are engineered as unrestricted products of domain analysis. To match the domain as closely as possible, networks were typically not structurally restricted in any way (Mayo and Mitrovic, 2001).A major hurdle for such systems is to define conditional probabilities, the value on each arc, in the absence of data. These probabilities might be suggested by experts and then verified later while the network is in use. Another disadvantage is that expert-centered BBNs may include so many variables that they become unwieldy and difficult to evaluate effectively online.

2.4 Conditional Probability Tables in Bayesian Model

These values for conditional probability are defined and a conditional probability table (CPT) is calculated. This probability distribution defines one entry for each different combination of values that variables can jointly take. For a directed model, the CPT is defined at each node. If the variables are discrete, this can be represented as a table that lists the probability that the child node takes on each of its different values for each combination of values of its parents. Assume that a BBN represents whether students can understand subtraction. Consider that the tutor begins with no clear knowledge about whether the student understands subtraction: There is 50% probability that the student understands subtraction and 50% probability that she does not.

Figure 2-3 A simple Bayesian belief network with a conditional probability table (CPT)

Solve P_{20}	P(3C = F) P(B = T)	P(B = T)
F	.5	.5
T	.9	.1

All nodes in this problem are binary–that is, they have two possible values denoted by T (true) and F (false). Consider the following example in which the goal is to infer students acquisition of knowledge given her behavior (e.g., number of hints, time to respond, correctness, and so on). demonstrates that the event of solving a subtraction problem, P 20, ″ Solve P 20 ″ (SP 20 = T) has two possible correlations: either the ″ student understands 3-column subtraction ″ (3C = T) or the ″ student understands borrowing ″ (B = T) or both. The strength of this relationship is shown in the Table. Either the student understands three-column subtraction (3C = T) or she understands borrowing (B = T).

3 THE ADAPTIVE TEST BAYESIAN NETWORK MODEL

3.1 Skills Model

Figure 3-1 Elementary and Operational skill and Misconceptions

In the above model the orange nodes represents Elementary Skills. The Green Nodes represents Operational Skills. The red nodes represent Misconceptions.

Elementary Skills
Addition (5/4 + 3/2)
Subtraction (2/4- 1/3)
Multiplication (2/3 * 4/9)
Comparison (3/2 > 5/4)

Operational Skills
Finding Common Denominator (1/2, 2/3) = (3/6,4/6)
Cancelling Out Technique (4/6 = 2.2/2.3 = 2/3)
Conversion to Improper Fractions ($3\frac{1}{2} = \frac{7}{2}$)
Conversion to Mixed Numbers ($\frac{7}{2} = 3\frac{1}{2}$)

Misconceptions	
About addition (MAD)	(a/b) + (c/d) =(a+b)/(c+d)
About subtraction (MSB)	(a/b)-(c/d)= (a-b)/(c-d)
Multiplication 1 (MMT1)	(a/b)*(c/b)=(a*c)/b
Multiplication 2 (MMT2)	(a/b)*(c/b)=(a+c)/b*b
Multiplication 3 (MMT3)	(a/b)*(c/d)=(a*d)/(b*c)
Multiplication 4 (MMT4)	(a/b)*(c/d)=(a*c)/(b+d)
Mixed Conversion (MC)	–

3.2 The Complete Model

Figure 3-2 The complete Student Model for the Topic of Fractions

The above figure represents the complete model for the Adaptive Test. Here we can see that the questions are connected to the respective skill set the question requires. Each of these arcs between the nodes represents a conditional probability. The Blue box at the bottom of the Model represents the overall knowledge of the student in the field.

3.3 The Question Set

Below are the questions that we used along with two options for each question. The last column shows the concepts used in each of the questions.

Question	Option 1	Option 2	Skills used
(14/17)>(18/25)	True	False	comparison
18/108	1/7	1/6	cancelling out
(15/19)-(19/38)	17/38	11/38	cancelling out, subtraction
((4/17)+(3/34)-(4/68))*(68/17)	7/17	8/17	cancelling out, subtraction, addition
87/11	7+10/11	1+8/11	CIM, MC
8/15*9/15	72/225	72/15	cancelling out, multiplication
((16/80)-((18/90)*(70/14))+(5/25))	-3/5	3/5	cancelling out, subtraction, addition , multiplication
((18/72)*(48/12))<(17/51)	True	False	cancelling out, multiplication, comparison, Subtraction
99/13	7+8/13	8+3/12	CIM, MC
v11/1331	1/121	1/11	cancelling out

3.4 Defining Each Question

In our example we have taken a total of five questions. Each question first must be represented in the Bayesian Network. When Representing the Network we need to connect the appropriate skill nodes with the question node. In the figure below we can see the Network model for question number 7. Since question number 7 uses the cancelling out, multiplication, Comparison and subtraction skills we can see that we have connected arcs for each of these skills. Second we need define the condition probability table for each question. The Conditional probability table for question number 7 is given in the figure below.

3-3 Network Model for Question no 7.

Cancelling Out	Knows								Doesn't Know							
Subtraction	Knows				Doesn't Know				Knows				Doesn't Know			
Multiplication	Knows		Doesn't Know		Knows		Doesn't Know		Knows		Doesn't Know		Knows		Doesn't Know	
Addition	Knows	Doesn't Know	Knows	Doesn't Know	Knows	Doesn't Know	Knows	Doesn't Know	Knows	Doesn't Know	Knows	Doesn't Know	Knows	Doesn't Know	Knows	Doesn't Know
Correct	1	0.97	0.92	0.82	0.91	0.73	0.63	0.42	0.78	0.72	0.67	0.39	0.53	0.35	0.28	0.12
Incorrect	0	0.03	0.08	0.18	0.09	0.27	0.37	0.58	0.22	0.28	0.33	0.61	0.47	0.65	0.72	0.88

3-4 Conditional Probability distribution for question no 7

4 ITEM SELECTION FOR ADAPTIVE TESTING

The Objective of Adaptive Tests are to gain as much information of the Student knowledge as possible. In the Our Adaptive Test Model we saw that it consisted of 10 questions. The questions are given to the student as such that the total information that could be gained is maximum. To select the questions we used the following Heuristics.

= log (+ 1 +) ..... (1)

Figure 4-1 The State Space for all possible strategy of Item selection. The above diagram is AND/OR graph.

The entropy for each skill is calculated according to the formula.

= ∈ ( ) x log ( ) …….(2)

Two Entropies are calculated for each question. The first Entropy (sum of entropies of all skills) is calculated after the evidence of the question is set as correct and the second entropy is

calculated after the evidence has been set to be incorrect. Now the final entropy for each question is calculates as

= +

……..(3)

( )

Here is the Probability of the question being correct and is the probability of the question being incorrect.

4.1 Psuedo code for entropy calculation

Below is the Pseudo code to calculate the entropy for each question.

```
FIND_ENTROPY(i)
    Set_Evidence[i] ="CORRECT"
    UPDATE_NETWORK
    S1 ← CalculateEntropy //according to equation 2
    Set_Evidence[i]="INCORRECT"
    UPDATE_NETWORK
    S2 ← CalculateEntropy //according to equation 2
    E← FindEntropy(S1,S2)  //according to equation 3
```

After the Entropy has been calculated the heuristics for each question is calculated according to equation 1. The question having the minimum entropy is selected as the next question.

5 APPENDIX

5.1 Screen Shots

Figure 5-1The First Question

Figure 5-2The test after few questions. We can see the knowledge Change.