International Journal of Approximate Reasoning 50 (2009) 450–460

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Bayesian networks: A teacher's view

Russell G. Almond *,[1], Valerie J. Shute [2], Jody S. Underwood, Juan-Diego Zapata-Rivera

Florida State University, ETS, Princeton, NJ, United States

ARTICLE INFO

Article history:
Received 1 February 2007
Received in revised form 29 October 2007
Accepted 21 April 2008
Available online 27 June 2008

Keywords:
Bayesian networks
Computer graphics
Probabilities
Aggregation

ABSTRACT

Teachers viewing Bayesian network-based proficiency estimates from a classroom full of students face a different problem from a tutor looking at one student at a time. Fortunately, individual proficiency estimates can be aggregated into classroom and other group estimates through sums and averages. This paper explores a few graphical representations for group-level inferences from a Bayesian network.

Published by Elsevier Inc.

1. Teachers' questions

Bayesian networks are becoming an increasingly popular way of representing the state of a student's knowledge, skills, or abilities, especially in intelligent learning environments (for example, ACED [1]). The display capability of most Bayesian network software is designed to work with one individual at a time. A teacher, however, is typically concerned with making inferences about a classroom full of students. This paper looks at the problem of making inferences about groups of individuals using the same Bayesian network.

Suppose that a teacher has 20–30 students who have taken an assessment which is scored using a Bayesian network. For each student, the teacher has a Bayesian network over a collection of *proficiency variables* which represents our best estimate of the student's state of proficiency. There are a number of questions the teacher might want to ask:

- How is the class (or school or district) doing overall? How many students are meeting or exceeding the curriculum objectives (standards)?
- Which students are not meeting the objectives? Which students are on the cusp of meeting the objectives?
- How are the students doing on each of the individual standards and skills (sub-proficiencies)?
- How does this class compare to other similar classes? Other classes in the same school or district? Classes in other districts with similar characteristics (where similar characteristics will depend on the purpose)?
- How do previously identified groups within the classroom differ?

* Corresponding author.
E-mail addresses: ralmond@ets.org (R.G. Almond), vshute@fsu.edu (V.J. Shute), junderwood@ets.org (J.S. Underwood), dzapata@ets.org (J.-D. Zapata-Rivera).
URL: http://ralmond.net (R.G. Almond).

[1] Several people have contributed data, ideas and suggestions which have improved this paper, particularly, Aurora Graf and Eric Hansen.
[2] Val Shute is now at Florida State University.

0888-613X/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.ijar.2008.04.011

- What are the typical patterns of skill acquisition (this is a more advanced question that might be asked when doing curriculum design or professional development, but not for routine classroom teaching)?
- How can the students be grouped into clusters which require similar kinds of instruction?
- Are there individuals with atypical patterns that require special attention?
- How much credence should I put in the estimates from the Bayes nets relative to other sources of information?
- What should I teach next? What should I do differently?

The goal of this presentation is to explore some graphical representations which might start to answer those questions. We will do this using data collected from a prototype system called ACED [1].

2. ACED

ACED (Adaptive content with evidence-based diagnosis [1,2]) is a computer-based assessment-for-learning system covering the topic of sequences, appropriate for a course in middle school mathematics. ACED is an experimental prototype designed to explore: (a) the use of the Madigan and Almond [3] algorithm to select the next task in an assessment, (b) the use of targeted diagnostic feedback, and (c) the use of technological solutions to make the assessment accessible to students with visual disabilities.

Graf [4] describes the construction of the proficiency model—a collection of latent variables describing the student's proficiency with sequences. ACED spanned three sequence types—arithmetic, geometric and other recursive sequences—commonly taught in 8th grade, but only the geometric sequence model is described here. The model is expressed as a tree shaped Bayesian network with the proficiency variables given in Fig. 1. Each variable can take on one of three proficiency levels: low, medium, and high.

The model was constructed through expert (Graf) judgment about the correlation between the variables and their parents in the hierarchy. The variables were chosen to reflect how the geometric sequences were represented in the tasks. There were 63 tasks in the geometric sequence portion of the assessment. In the evidence model for each task, the task outcome (evaluated as right or wrong) was directly related to (had as a parent) a single proficiency variable.

ACED tasks are based on the National Council of Teachers of Mathematics standards which in turn form the basis of the standards of all 50 US states. Although firmly based on those standards, true alignment is difficult to achieve because (a) ACED has a finer level of detail in its (diagnostic) proficiency model than is found in most standards, and (b) all 50 states have set the cut point for "proficient" with the general category of sequences at different places. Thus, although performance on ACED should be strongly correlated with each state's standards, the medium proficiency level in ACED may be higher or lower than the proficient cut point set in any given state.

The data used in the graphs below comes from an evaluation of ACED [2]. It consists of data from 157 students who received the adaptive[3] version of ACED. Roughly half the students received diagnostic feedback designed to help them understand their mistakes; the remaining half had accuracy-only feedback. For this paper, we will ignore the evaluation component of the study (including pre- and post-test measurements) and focus on the data that can be used to produce a collection of representative scores that a teacher might see. Note that for these students, geometric sequences were not an explicit part of the curriculum, although some geometric sequence problems may have been taught as part of other topics in algebra.

3. Scores coming out of a Bayesian network

ACED scores student responses using the Bayesian network. The individual task outcome variables are entered as findings in task-specific nodes and the results are propagated through the proficiency model. After evidence from all tasks are entered, the posterior proficiency model gives our beliefs about the proficiency state for this particular student. Technically, any *statistic*—that is a functional[4] of that posterior distribution—can be used as a score. In practice, the marginal distributions for each of the proficiency variables in Fig. 1 were recorded for each participating student.

Let $S_{i0},\ldots,S_{ik}$ be the proficiency variables for student i, with S_{i0} representing the special overall proficiency variable (SOLVE GEOMETRIC PROBLEMS in ACED). Each variable can take on the values low, medium, and high. Given a body of evidence, $\mathbf{X}_i$ the Bayes net can efficiently calculate $p(S_{ik}|\mathbf{X}_i)$, the conditional distribution of S_{ik} given the observed outcomes. Four statistics of this marginal distribution are of particular interest.

Margin: The marginal distribution of the proficiency, $p(S_{ik}|\mathbf{X}_i)$. This has the disadvantage of being three numbers (summing to 1.0) so it is not compact to display or simple to interpret.

Cut: If one of the states has special meaning, e.g., students at the medium level or above are considered to be "proficient" on some set of standards, then the probability that the student is proficient, $P(S_{ik} \geqslant \text{medium}|\mathbf{X}_i)$, is just a sum of

[3] The adaptive version of ACED used the Bayes net to select the item sequence based on the pattern of scores observed so far, and hence Bayes net scores were more readily available [2].

[4] A *functional* is an operator which maps a function (e.g., a probability distribution) to a scalar quantity.