Use of Bayesian Models, Markov Models and Data Mining in Intelligent Tutoring Systems

Spyridon Psychogios, Greta Kaufeld and Andy Russell

March 5, 2014

1 Bayesian Modelling

1.1 Working Definition

"The Bayesian network is a graphical modeling tool for specifying probability distributions" (Darwiche, 2009, 53).

1.2 Where

Student Modelling, for example in ANDES (Conati, Gertner, VanLehn, & Druzdzel, 1997): A Bayesian network is used to do long-term assessment of the student's domain knowledge, plan recognition - inferring the most likely strategy the student is using to solve a problem - and predictions of students' goals and actions during problem solving. Support can then be tailored to the students' needs based on the information in the student model.

1.3 How

Conati et al. (1997) start out with a so-called *problem solution graph*, which is a hierarchically structured dependency network including all acceptable solutions to problems as well as abstract plans for generating those solutions. The solution graph is then converted into a Bayesian network by annotating all the top level nodes with prior probabilities, while all other nodes are annotated with conditional probabilities relating to their parent nodes. While students take actions, ANDES determines which nodes in the network the action corresponds to and the probabilities in the network are updated.

1.4 Why

Bayesian modelling is used in order to handle uncertainty regarding *knowledge tracing*, that is, inferring what domain knowledge a student has, as well as *plan recognition*, that is, inferring what goals (s)he is currently trying to achieve, from the actions (s)he takes (Conati, Gertner, & Van Lehn, 2002).

2 Markov Models

2.1 Working Definition

Markov Models are probabilistic Finite State Machines that consist of a set of states connected by transitions. As suggested by the Markov property, each state depends only on the previous n ones (Renals & Hain, 2010). Hidden Markov Models are named that way because their state sequence is not observable: only the output is. The include the following parameters: Transition probabilities between states, observation probabilities, and initial probabilities for each state (Jeong et al., 2008).

2.2 Where

For example, in Betty's Brain (Jeong et al., 2008), HMMs are used to determine students' pattern of activities: What strategies do students use in order to create concept maps and teach Betty?

See also D'Mello and Graesser (2010) and Beal, Mitra, and Cohen (2007) for more applications of HMMs in ITSs.

2.3 How

For Jeong et al. (2008), HMM sequences of states correspond to students' learning behavoir patterns: States can correspond to single tasks, such as editing the concept map, or multiple ones, such as quizzing Betty and accessing the reading resources. Transition probabilities model the probability that a student transits from one state to the next or remains in the current one; stationary probabilities give the probability of a student being in the current state. HMMs are trained from log files which were recorded during study sessions with the tutoring system. Patterns could be extracted from these HMMs.

2.4 Why

Because HMMs go beyond simply "counting" occurrences and give the researcher the opportunity to extract information about coherence between different phenomena (Jeong et al., 2008). They can provide a more global view of the learning process (Jeong et al., 2008).

References

Beal, C., Mitra, S., & Cohen, P. (2007). Modeling learning patterns of students with a tutoring system using hidden markov models. In *Proceedings of the 13th international conference on artificial intelligence in education* (p. 238-245).

Conati, C., Gertner, A., & Van Lehn, K. (2002). Using bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction*, *12*, 371-417.

Conati, C., Gertner, A. S., VanLehn, K., & Druzdzel, M. J. (1997). On-line student modeling for coached problem solving using bayesian networks. In A. Jameson, C. Paris, & C. Tasso (Eds.), *User modeling: Proceedings of the sixth international conference.* Vienna, New York: Springer Wien New York.

Darwiche, A. (2009). *Modeling and reasoning with bayesian networks.* Cambridge University Press.

D'Mello, S., & Graesser, A. (2010). Modeling cognitive-affective dynamics with hidden markov models. In R. Catrambone & S. Ohlsson (Eds.), *Proceedings of the 32nd annual cognitive science society.* Cognitive Science Society.

Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. (2008). Using hidden markov models to characterize student behaviors in learning-by-teaching environments. In *Lecture notes in computer science* (Vol. 5091, p. 614-625). Springer.

Renals, S., & Hain, T. (2010). Speech recognition. In *The handbook of computational linguistics and natural language processing.* Wiley-Blackwell.