422 - Reading 2 Questions

Student Tendencies

1. [Rephrased] When modelling a student's reading time, the paper assumes that a longer reading time suggests a higher probability that the student has self-explained. Will this assumption accurately assign credit especially in cases when the student is distracted, when the natural reading speed of the student is considerably slower than the average, or when the student's first language isn't English? (**Asked ~7 times)**
2. [Rephrased] How will the system deal with situations when the students are unable to describe their thinking, make calculation errors, or write alternate solutions, but they understand the rule or knowledge well? Is there a way to model errors arising due to such situations?(**Asked ~5 times)**
3. The paper discusses how the Bayesian network is capable of modelling the probability of a student guessing the answer to a problem, but it does not take into account the possibility of a student cheating. How could cheat detection be modelled into the Bayesian network to provide an even better assessment of a student? **(Asked ~2 times)**
4. Shouldn't the leaky probability be student specific, because some student's are more likely to ask external sources for help or often guess? **(Asked ~2 times)**
5. The researchers found significant improvement of students using Andes (and other ITS) compared to students undergoing normal instruction alone. Is it possible that students improved not because of ITS use, but because they were simply practiced more?
6. Different students have different speeds of learning and start off with different domain knowledge. Is there an easy way to gather initial prior probabilities of mastery for a first time user of the program?
7. If students with learning disabilities were to use this program, they may be prone to be random. Would this break the system since the program would not be able to make a prediction?
8. Could retaining information about common mistakes be used to help the system coach the students in a more meaningful way?
9. What if there are external factors in a student's life that may affect their learning capability, such as mood, trauma, motivation, and procrastination?
10. What are the students' objective/incentive in using the system? Do they actually want to learn, or just to get it correctly and get it over with when they use the system?
11. Does the system have a preset template network for the type of problem solver you are, and then assigns a student to it after a small sample? If so, what if the system mistakenly assigns you to a problem solving style that is not how you like to do problems, wouldn't the system be more confusing as it gives you hints at the wrong time?

Methodology

1. What are some other approaches towards building intelligent tutoring systems? (**Asked ~3 times)**
2. Is there a trade-off in implementing the ITS like this?
3. Is it possible to create an ITS that both teaches and learns from feedback? Or a similar system? How would that scale in terms of networks, and be implemented?
4. How should trade-offs be implemented in such as less knowledge assessment accuracy versus more responsive and generalized hints and solutions in problems that have more linear solution paths.
5. How would the plan recognition be improved, if hints were not relevant to what users were thinking about at the time they asked for help?
6. The Andes system tries to help a student learn material by looking at a student's past actions and experience. Is it better to teach a student new material by using old concepts/rules or will using new concepts/rules make a student's understanding of the material even better? This is very similar to the exploit/explore concept for MDPs.
7. How much extra effort is required on the part of the teacher to encode the information if they wish to add extra questions? Is the tradeoff in question quality vs. quantity worth it?
8. it's stated that having one template for the problems is not feasible, nor is adding all by hand. What is the solution then?
9. Does it make sense to use the same rules for *solving* the problem versus *explaining* the problem? The example explainer uses these same rules, but the granularity of said rules are very small and may be too algorithmic for a student to find useful (in some ways, this is like how a computer algebra system solves derivatives; the steps tend to be needlessly systematic).
10. Is the trade-off between students freedom, and correct predictions justified?
11. Is there a way to overcome the assignment of credit problem without compromising the experience of the student?
12. Why did the evaluation with material covering larger scope ended up to be more successful than the one with smaller scope (1 grade increase vs 1/3 grade increase)? Does it imply that the system provides stronger conceptual understanding?
13. If the theory of hints used in section 3.2.2 was changed so that hints are capable to change the domain knowledge of students (either through rewording or increasing the presence of explanations), how would this affect how hint nodes are used in the belief network? This is brought up only because in scenarios where there is more self-learning (either systematically through the nature of the course or variability in student learning) it may be useful to have the hints in a ITS be a source of learning also, and modelling this may be useful.

Data Collection

1. The solution requires manually collecting domain knowledge and questions - is there a way to automate the process or more easily build domain knowledge, for example to apply this solution to a different curriculum than the targeted Newtonian physics?
2. Evidently, collecting more information about students' actions and inferences helps Andes better tailor to the student. At what point does one stop pursuing data collection in favour of maintaining student interest/initiative?

Model

1. How much knowledge would a programmer need about the underlying Bayesian Network to add a new question to an application based on this model?
2. The modeling seems quite complicated, how are all the scenarios of the model determined?
3. Is it possible that the model would be able to answer the question itself (based on the same text that the student is using), or does it just rely on the 'coded' versions of the questions in tutoring the student?
4. If Andes had been used to evaluate students using problems that it decided were more suitable to each individual student rather than use the same problems for everyone, how would the assignment of credit have changed considering that students would now be answering questions that are specifically meant to challenge them based on their previous inputs?
5. Can student model and plan recognition be combined into a single network?
6. In terms of the leaky-problem, does this apply only to features of the question that are independent of rules? I.e, for some question is it seen as more or less likely that a student will guess which physical law to apply to a question?
7. Why is the leaky-OR gate more likely to lead to success as a parameter indication?
8. How exactly does the model generate the solution graph of a specific problem given a proposition and a goal, wouldn't different problems need the author to specifically tell the system what the exact rule-context is for that problem?
9. Why did they base their inferences on the updated conditional probabilities of the nodes, when in lectures we saw that the inferences should be calculated conditioning the node to the entire evidence somehow (like with MCMC)?

Applicability of model to other subjects/ Extensions

1. [Rephrased] How easy is it to adapt this model for use in other subjects that are less mathematical, or are more subjective in nature, or don't have well defined rules, or where problems can have multiple solutions (like language, philosophy, business, chemistry etc) **(Asked ~30 times)**

2. Has this been tried out in simpler courses, such as primary school education? It seems like Bayesian network will scale better in easier courses such as primary school math classes as questions and solutions are simpler **(Asked ~5 times)**
3. Would it be possible extend this Bayesian network model to teach a classroom instead of an individual student? **(Asked ~5 times)**
4. This technology is really fascinating, I'm curious to know why it is not more common place. Why was I not able to have an ITS help me learn while I was taking Physics? (**Asked ~3 times)**
5. Is it possible to apply this system to higher levels of education, or is there a ceiling in terms of the scope/difficulty of content it covers? (**Asked ~2 times)**
6. Can the same technique be applied to more sophisticated fields/levels? Such as Advanced Physics taught in university. (**Asked ~2 times)**
7. Where could this technique be applied to improve our modern learning models?
8. How could the same idea be integrated with existing probabilistic models to improve accuracy

Work since Paper was written

1. [Rephrased] How can modern natural language processing be used in this implementation to improve performance? (**Asked ~8 times)**
2. [Rephrased] Have there been any improvements in Andes or advances in the implementations of ITS since 2002? Is there a more advanced tutoring system based on a similar approach today? (**Asked ~5 times)**
3. Have the network size and resource problems become less significant over time? Was this just a limitation of computing power when the paper was written? Were faster update algorithms developed? (**Asked ~2 times)**
4. I understand this paper was published in 2002 and I was wondering whether these models are now widely used?
5. Where is this work going in the future? Can this technology be used to quickly identify which concepts a student lacks and give highly specific tutoring on a certain aspect of a subject?
6. With modern data storage capabilities and more advanced machine learning capabilities, does Andes framework more accurate today than it was when the article was written?

Scalability

1. How does this system respond to scaling up, i.e. handling bigger problem size? What is the run time with respect to problem size? **(Asked ~8 times)**
2. How much time would it take for the system to update and provide the step-by-step help for a particular problem? (**Asked ~3 times)**

3. How did they deal with some of the scaling issues that weren't solved by using Bayesian Networks?
4. Since many actions and identical rules may be used to solve a physics problem, the Bayesian network can be quite large and interconnected. How do we update such networks in real time?
5. Can the time complexity of the approximate inference model be reduced?
6. How will the Bayesian scale with more problems and complex actions with a more competitive pre-condition knowledge base leading the credit assignment to each individual knowledge become more difficult?

Bayesian Network

1. [Rephrased] As one of the goals of this project is to estimate the state (plan recognition) based on actions, can MDPs/POMDPs be used instead of Bayesian networks? If so, how do these methods compare to Bayesian Networks? (**Asked ~10 times)**
2. [Rephrased] What was the intuition behind using Bayesian Networks over other methods to solve this problem? (**Asked ~6 times)**
3. Can Bayesian networks be used in other AI systems (ex. customer service, help desk, etc.)? (**Asked ~3 times)**
4. Will Bayesian networks work well in building most models that includes uncertainty problems? (**Asked ~2 times)**
5. The authors talk about preserving and passing information from a replaced part of the Bayesian network to another subsequent corresponding part. Could we repeat this indefinitely? If so, why do we need MDP? Is not there some trade-off if we use a Bayesian network, for instance some aspects of the domain not being reflected well thereby limiting accuracy?
6. Rules in the form of if-then statement used to generate a solution graph seem as if logical in nature, i.e. certain. Given this, how can we give a probabilistic interpretation when a solution graph is converted to part of a Bayesian model? What is the logical or mathematical basis for this?
7. Wouldn't testing each individual policy (such as guessing, hints, credit etc.) be useful in terms of creating a more robust Bayesian network to handle this problem?
8. Does a student need to every type of question for the Bayesian network to be accurate in predicting student mastery?
9. What makes Physics a suitable subject to be modeled with the dynamic Bnet approach?

Criticisms/Suggestions for Experiments

1. How would you validate this solution in a method other than empirical testing? (**Asked ~4 times)**

2. The evaluation of the solution compares students who used the SE-Coach amongst each other. But I am curious to know: How did students who used the SE-Coach compare to those who did not? Did using the SE-Coach yield significantly higher scores on the posttest?
3. Are there any proven evidence that students have their learning made easier and grades scored higher using this program?
4. Has the AI been tested on a subject with more clear-cut solutions? (IE, biology or medical students).
5. Were further studies conducted, or was it just a specific group of children they were working with. Perhaps these students, had more knowledge and access to certain materials that helped them along with the Andes' system?
6. Why did the author stick to only one intelligent tutoring system (mainly the Andes system)? Why not try applying the analysis to more existing systems and see how they differ?
7. Since learning is such a subjective term, were methods other than comparing student letter grades considered when evaluating user performance?
8. How would the results change for older/younger students?
9. The amount of time that students were allowed to use Andes was not addressed. This could have meant that students who used Andes could have been exposed to help more than other students who did not. Would controlling the amount of time being exposed to Andes to equal amount of time of help that other students have received impact the average grade?
10. In practice, does a system like this really encourage the development of subject mastery, or does it simply encourage students to develop pattern matching on problems?

Probabilities/Parameters

1. How were the initial conditional probabilities decided on, and how can it be proven that these accurately model the correlations between student knowledge, goals, and behaviour? (**Asked ~2 times)**
2. In section 2.2.2.1, the leaky probability beta is described as being the probability that somebody guesses a rule application without all of the prior evidence. How would this value be determined? **(Asked ~3 times)**
3. What is the best way to determine the initial parameters?
4. How to define the value of marginal probability in the network?
5. Can you clarify how were the new probabilities calculated after an action was taken
6. Does estimated probability of mastery degrade over time so that previously "mastered" topics are still studied once in a while?
7. How do you determine the exact value of noise parameter
8. How is the value of the (alpha) constant determined for a context-rule node in the network?

Limitations/Criticisms

1. It seems like there is a lot of work that goes into adding a single question to the problem set. Barring the ease of changing out the numbers or other features of a question, is it true that there is a limited ability to add new questions? (**Asked ~3 times)**
2. Is this proposed solution actually solving the problem to an extent or just being more specific with the cases that the Bayesian network will handle?
3. How can you guarantee that no question or example will automatically generate a subnetwork so complex that calculating the prior probability in the ‘rollup’ function becomes impractically slow?
4. When multiple strategies are possible to solving the problem, why isn't the "easiest" strategy (though that could be hard to quantify--maybe use number of steps) weighted more heavily, since a student would be more likely to use that solution?
5. If a student uses many hints to arrive at a correct solution, how can we be sure they actually learned the material and rules?
6. Andes tries to simulate the traditional *paper-and-pen* homework experience and gives feedback. Isn't the traditional learning method more convenient than a tutoring system? For example, typing a formula is slower than writing one.
7. Do the authors have a plan or idea of how to fix the uncertainty problem that students have the initiative and “head start” on the AI?
8. How are we sure that the student is struggling with the question rather than the student had trouble inferring what the question is looking for?

Andes System

1. Instead of requiring students to display more of their thinking, what other ways would improve Andes accuracy in student modelling?
2. Andes accepts drawings of vector diagrams on its interface. Would it be able to accept drawings of equations as well instead of having to type them, as they may be easier to input that way and could thus, possibly increase student acceptable?
3. Andes makes the distinction between rules and context rules. Is the context rule-related (context of rule with other rules), or application-related (context of rule in problem scenarios)?
4. How does Andes determine and account for the level of mastery of skills required outside of the given course? i.e. mathematics.
5. How many average questions does it take for andes to predict the student behaviour?
6. User interfaces in general have evolved quite a bit since the last time Andes was evaluated. Would qualitatively better UX improve the evaluation results?

7. Can the design of the interface affect the resultant statistic data generated from the user's interactions? Different people learn different ways. It seems like the Andes program heavily involves visual interaction. What about audio?
8. If explaining the material was a built in feature of the system, would it also improve the performance of the system as the student's knowledge base would be more easily accessible by Andes?
9. Because Andes does not require students to explicitly state all their reasoning for an answer it limits the amount of information the student model has for predicting. Why don't they prompt students to explicitly state all their steps in arriving at an answer and even if they did, would it be enough to generate a more accurate prediction?

Clarity Issues

1. It is unclear to me how the automated solution generation works?
2. What is a "SE-Coach"? Is this just an interface for human instructors to monitor student responses?
3. I'm a bit confused about how different nodes interact with each other, especially the strategy nodes and proposition nodes. I hope we can go through a concrete example that has all types of nodes.
4. How does latency data helps the coach determine probabilistically when the student constructed a self-explanation mentally, without entering it using the menu-based tools.
5. How do you chain the events/probabilities of what students are going to answer? There are so many possibilities what a student can answer.
6. What is considered as a 'state' in their Bayes Net?
7. In the prior probability calibration for rule nodes, why were the "non-tested" rules initialized to 0.5, as opposed to, say, the average of the rules set from the multiple-choice test data?
8. Why was conditional probability evaluated for a proposition node and not the joint probability distribution?
9. For the conditional probability table of RA = T given its parents, why shouldn't P(RA = T|at least one parent = F) not be equal to 0 but instead be a value based on how many of its parents are true.

Possible Modifications

1. Would a neural network possibly be beneficial in regards to the last paragraph of the paper, which says that a model that can better determine when a tutorial/exploratory approach is better would overall be richer? (**Asked ~3 times)**

2. Is it possible to add some machine learning ideas to this system so that the system could start to learn students' actions over time? (**Asked ~3 times)**
3. The system seems to not be explicitly leveraging external indicators of student learning such as midterms, quizzes and finals, which makes sense since it is a ITS which is isolated only for problem solving practice. But, if we wanted to import external data such as a midterm marks (assuming all the questions are compatible with Andes), can this be done by just altering the probability of mastery for each of the rules in question for the student? (**Asked ~2 times)**
4. Would it be possible to selectively present questions to the student to gather evidence about particular nodes in the network? For example, if we know the student has high understanding of CR1 and we're confidence in that observation, we can show them problems that would have them invoke CR2 (sibling node of CR1) in order to determine whether they have a general knowledge of the parent rule node. How would something like this be formulated?
5. Could we simplify the network by reducing some of the machine learning part into user actions. For example, instead of deciding which students need self-explaining and only display self-explanation to those students when they try to leave the example, we could just ask all students if they want to go through self-explaining or not. This way, other students also have the option to review and solidify the knowledge.
6. Will having pre-determined student profiles that categorized and estimate student's ability be a less complex solution?
7. The paper mentions the difference between a "constrained" and an "open" approach for the learner. Could Andes be updated such that Andes is able to decide when one is appropriate over the other? Perhaps based on the type of domain, or maybe inferred attributes of the learner themselves.
8. They give immediate feedback for student's actions, and they allow the student to ask for hint whenever she wants. Wouldn't it be better to allow a condition of when to refuse to help the student, or to make some nodes that doesn't lead to immediate feedback, to let them re-think by their own if their procedures makes sense, as happen when being taught by a professor? Maybe the student performed better at the end in standard problems, but they could learn how to apply the rules without being aware of their meanings, like in RL.
9. Is there possibly dependencies between context rules that could be used to improve the network?
10. Could other statistical theories/models been applied into the model so that it will perform better in dealing with uncertainty?