How to orientate arcs in a Bayesian network based student model?

Mathieu Hibou*
Jean-Marc Labat**
*CRIP5/AIDA Université Paris 5 René Descartes
**LIP6/AIDA Université Paris 6 Pierre et Marie Curie
mathieu.hibou@math-info.univ-paris5.fr
jean-marc.labat@upmc.fr

Abstract
Bayesian networks have been successfully used for student modeling in many systems. In this paper we address the problem of Bayesian network structure construction and more particularly that of arc orientation. We think that, in the case of cognitive task modeling, the traditional causal interpretation of arc orientation is not adequate. Instead, we use the information flow to provide a systematic a priori analysis of the conditional dependencies between variables in the case of overlay student models. Finally, we explain why we think that different Bayesian networks should be taken into account and how this could be done.

1. Introduction
Bayesian networks have been successfully used for student modeling in many systems whose purposes and functionalities greatly vary. Some of them are only dedicated to evaluating the student's competence [11], while others make it possible to recognize the plan followed by the learner [7]. Moreover, they have been implemented in different frameworks: in “traditional” ITS, in Open Learning Environment [2] or as a basis for inspectable models [18]. This variety of uses advocates strongly for Bayesian networks in the implementation of student models.

A Bayesian network [13] is a set \([X_1, X_2, \ldots, X_n]\) of random variables associated with a directed acyclic graph so that the nodes of the graph are in a one-to-one correspondence with them, the joint probability distribution being given by:

\[
P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | pa(X_i)),
\]

where \(pa(X_i)\) represents the parents of \(X_i\) in the graph. The structure of the graph is in fact a translation of conditional dependencies between random variables, and thus it has an influence on information circulation. Once the network is specified, inferences can be made. An inference in a Bayesian network is an update of the probabilities, that is to say the calculation of the posterior probabilities of the unobserved variables when any evidence is observed.

In this paper we address the problem of Bayesian network structure construction and more particularly that of arc orientation. Our key idea is that, in the first section, we present very briefly the different ways of building Bayesian networks, then we set up the general background our considerations apply to before analyzing the arcs orientation between different types of nodes.

2. Bayesian Networks
The complete specification of a Bayesian network requires the definition of both its structure (a directed acyclic graph) and parameters (the probability tables). This can be done either by experts’ knowledge elicitation, by using machine learning, or even by mixing both approaches.

In the case of knowledge elicitation, the definition of the network structure begins with the collection of the possible nodes. A distinction is made between informational variables (the ones that can be observed) and hypothetical ones. The presence of an arc is the expression of the potential influence between two variables. Usually arc orientation is analyzed in terms of causality: the existence of an arc from A to B means that A is one of the causes of B. In section 3 we discuss the difficulty of such an interpretation in the case of learner modeling. The parameters are determined in an approximative way, using qualitative information.

There are two different manners to learn a network from data: the statistical one and the Bayesian one. In the case of statistical learning, the maximum likelihood method is used, whereas in the case of Bayesian learning it is the maximum a posteriori estimate that serves as a
criterion for learning. The Bayesian learning consists in choosing the model that has the highest probability given the data. As explained in the next section we focus on the expert elicitation methods and we invite the reader to consult [3] for more information on that topic.

3. Student modeling

We do not aim at examining the problem of arc orientation as a whole and therefore we focus on the relatively frequent case of the use of Bayesian networks to represent overlay [14] student models. Andes [7] is typical of the kind of models we focus on: these are models where part of the knowledge on the student is totally unobservable. In the case of Andes, it is impossible to observe whether or not a student masters a physics rule.

Additionally to that condition, we consider our model to be split into 3 types of nodes, knowledge nodes, know-how nodes and item nodes. The first and the second types of nodes constitute what is often referred to as the *domain part* of the graph while the third type nodes are the observable bricks of the *task part* of the graph (as in [2], [4] or [9]). Because item nodes are observable, there is a relation between them and the interaction of the student with the interface. These items can be the evidence of many different actions such as clicking on a number in Prime Climb [9], or entering a newly deducted fact in the Andes interface.

We take into consideration items that reflect knowledge operationalizations (typically maths exercises or physics problems). Therefore it is reasonable for item nodes to be linked only to know-how nodes because know-hows are the operationalized knowledge.

Factorization nodes in Prime Climb are good examples of know-hows. The kind of model we consider just as *Prime Climb* does not have to necessarily integrate knowledge nodes. In many systems, the difference between theoretical knowledge and practical know-hows is not made ([2], [4] and [7]). However we think that this distinction is important in terms of topological constraints on the network, as shown in sections 5 and 6.

We focus on the expert elicitation of the structure, because in the case of unobservable knowledge and know-hows it is really hard, if not impossible, to properly set it only on the bases of the data observed. And even in Capit [10], or in the system described in [17] where the model structures are built according to the data, there are experts' interventions at least in order to fine tune the graphs.

Is a student good at mathematics because s/he is able to solve many different problems, or is s/he able to solve many different problems because s/he is good at mathematics? As already mentioned, we believe that in the case of student models, the causal interpretation of the arcs is irrelevant. In many cases the arcs are oriented with no strong analysis and sometimes changed *a posteriori* because of the inefficiency of the model [9]. Therefore, we propose to analyze the flow of information [15] in the network in order to deal with this orientation problem. Given a set of observed nodes $O$, the flow of information between two nodes is blocked by two types of nodes:

- the observed nodes, except the one with converging arcs,
- the unobserved nodes with converging arcs, except the ascendants of one of the nodes in $O$.

In the next sections we analyze arc orientations between the different node types presented before. We first begin by exposing our considerations and then we examine examples in order to confirm our analysis.

4. Arcs between item and know-how nodes

Let us consider two items that need the same know-how to be applied (the corresponding nodes are called $I_1$, $I_2$ and $KH$). Either $KH$ is $I_1$'s and $I_2$'s father, or $KH$ is their common son (figure 1). In the second case $I_1$ and $I_2$ are independent variables (no flow of information from $I_1$ to $I_2$ if $I_1$ is observed) because $KH$ is unobservable. This does not seem to be reasonable: the two items are the application of the same know-how, consequently any evidence on one of them should give information on the other. The result is that arcs have to be oriented from the know-how node to the item node.

Another point strengthens our position. If we consider an item $I$ that is the expression of two know-hows $KH_1$ and $KH_2$, either $KH_1$ and $KH_2$ are brothers, or $I$ is their common son (figure 2). In the first case, $KH_1$ and $KH_2$ are conditionally independent given $I$, and...
which also advocates for the second case to be used, that is to say orientation from the know-how to the item.

In the literature this is exactly what happens for examples in the systems described in [7] and [17]. In the case of the educational game Prime Climb [9] there is even an evolution of the student model (for efficiency reasons) corresponding to our analysis. In the first structure the arcs were oriented from the items (click nodes) to the know-hows (factorization nodes), the principle being that you have to click on a number X if and only if it does not share any factor with the number K on which your partner is (figure 3). After an a posteriori analysis close to our second point (the factorization nodes should not be independent given the click node), one of the modifications made to get a more efficient model has been to draw the arcs from the factorization nodes to the click nodes. In order to get valid guidelines for Bayesian network construction, we strongly advocate for an a priori analysis.

As an example, let us consider that K is Pythagoras' theorem (presented as an equivalence), and that KH1 and KH2 model respectively the ability to use it to calculate lengths or to prove the existence of a right angle. Basically what is the knowledge in that theorem? It tells us about the relationship between the existence of a right angle in a triangle and a certain relation between the square of its sides lengths, the main difficulty being to remember the precise form of this relation. A valid use of KH1 should lead to believe that KH2 is well used too, because it means that the student knows the form of the relation between the square of the triangle side lengths, and hence s/he should be able to use it to prove that a triangle is right-angled.

Therefore, arcs must be oriented from knowledge to know-how nodes.

6. Arcs between knowledge nodes

In section 5 example, we implicitly suppose that the students are at a level where Pythagoras' theorem is presented as an equivalence. Earlier in their curriculum, when the theorem is presented in a twofold way (the theorem and its reciprocal), the two know-hows mentioned above would not have been linked to the same knowledge, but to two different ones, one for each implication of the equivalence. If we want to model the general knowledge about Pythagoras' theorem we have to add a new knowledge node, linked to the node «direct theorem» and «reciprocal theorem». Therefore, we analyze the orientations of these arcs between knowledge nodes in this section.

There are mainly two options for the arc orientation between knowledge nodes: from topic to sub-topic (figure 5a), or the reverse (figure 5b). In the second case the sub-topics (and the items testing them) are independent and they are dependent in the first case (again because the knowledge nodes are unobserved). In
Hydrive [11] and Andes [4] the first option is chosen: the arcs are oriented from general to more specific knowledge. About this choice in Andes, the authors note that it implies an (eventually testable) hypothesis on transfer between knowledge on context rules depending on the same physics rule. We agree with that remark: arcs orientation between knowledge nodes cannot be decided in general, as opposed to the cases analyzed in sections 4 and 5. We think that the choice made in Andes is also due to the unclear status of context rule nodes (knowledge or know-hows), hence in favor of a solution compatible with section 5 (the rules are clearly theoretical knowledge in Andes).

Orientating the arcs from topic to sub-topics comes to considering the sub-topics as an emanation of the topic in that the general knowledge is somehow determinant for the more specific ones, whereas the other orientation seems to correspond more to a whole-part relationship [7].

Our opinion is that the choice of one of the orientations depends mainly on the nature of the general knowledge and on the level of expertise of the student. The more composite the general knowledge is, the more it advocates for an orientation from sub-topics to topic.

A similar orientation seems to be also adequate in the case of novice students. We use again Pythagoras' theorem as an example. The year the students are taught the reciprocal theorem for the first time (the direct version being taught the year before) it could be reasonable to orientate the arcs from the direct theorem and reciprocal theorem nodes to the Pythagoras node, because at that stage of the learning process the direct version can be mastered whereas the reciprocal one cannot. In fact these considerations about the variation of the graph structure according to the expertise level of the student leads us to search in the direction of multi-network models [5] (see section 8).

7. Arcs between know-how nodes

In this section we analyze the dependency relationships that result from the arcs between know-hows, regardless of the ones resulting from the upper part of the graph.

Figure 6.

Let us consider three know-hows, KH1, KH2 and KH3, and two items I1 and I2, respectively associated with KH1 and KH2. We suppose that KH3 is a compound of KH1 and KH2. For example, KH1 could be “calculate with powers”, KH2 “calculate with square roots” and KH3 “calculate with both square roots and powers”. I1 and I2 are independent if and only if the arcs are orientated from the less to the more composite nodes (figure 6).

The situation could also have been modeled by directly linking KH1 and KH2 to an item I3 (testing the compound know-how), the arcs being drawn from the know-how nodes to the item node (figure 7), according to the considerations of section 4. In this case if I3 is unobserved, I1 and I2 are independent. These two ways of modeling the situation must be equivalent (it is possible to use both of them for different situations, depending on the frequency of the different compound know-hows) and therefore the arcs need to be oriented from the know-how nodes to the compound ones.

Figure 7.

Again we insist on the fact that we do not take into consideration the knowledge part of the graph. If we come back to our example, that means that the items testing respectively the know-hows “calculate with powers” and “calculate with square roots” maybe dependent, but through the relationships between the knowledge “powers” and “square root” and not at the know-how level: it is the knowledge part of the graph that encodes the expertise level of the learner as explained in section 6.

Hence, contrary to the case of arcs between knowledge, it is possible to make a strong hypothesis on the arc orientation between know-hows: from the less to the more composite.

8. A multi-network model

The learner's level of expertise is not fixed. It varies through time and interactions with the system, and therefore the student model has to take these variations into account. As explained in section 6, the different levels of expertise are translated into the graph by different arc orientations between knowledge nodes.

Hence a Bayesian network based student model must be constituted of several different networks, with different structures. Moreover, the evolution of the student model's structure is not only a question of coherence: the structure of the Bayesian network gives a diagnosis on the learner's level of expertise.

In order to test and validate the use of such a multi-network model, we consider the formal structure of
Bayesian network mixture (or multi-nets) [5]. The principle is to consider different Bayesian networks and to associate to each of them a probability (the probability that represents the network frequency in the considered sample).

We use data from Pepite [8], a diagnosis system in mathematics and build, according to our prescriptions on arc orientation, different network structures reflecting the different possible expertise levels. We consider a global Maths node composed with two general fields of knowledge, Geometry and Algebra, each of them being the combination of several sub-topics (such as Fractions or Square Roots). Two choices of orientation for each relationship between these general fields and their sub-topics, and two choices for the link with the Maths node lead us to build 8 different network structures.

The coherence test of our model with virtual students is positive. For example, the network that fits the best a virtual student that masters all the topics in geometry and all of them except one in algebra is effectively one with the arcs oriented from Geometry to its sub-topics and from the sub-topics of Algebra to Algebra.

We will now continue our tests with real students' log files, in order to compare the single network model with the multi-network model.

9. Conclusion

In the context of overlay models constituted with three types of nodes, knowledge, know-how and item nodes, we have explained how the number of combinations of arc orientations can be reduced by analyzing the conditional dependences between variable.

In section 4 we provide strong evidence on the orientation from the know-how nodes to the item nodes. We explain in section 5 and 7 the reason why we think that the arcs have to be drawn respectively from the knowledge nodes to the know-how nodes and from the general know-hows to the particular ones.

As indicated in section 7, the orientation between knowledge nodes depends on the learner's expertise level, and because the expertise level varies from one student to another (even in the same class) models should take this into account.

According to our analyzes on the network structure variability, we have begun to test a framework (mixture of Bayesian networks) to see if instead of choosing an orientation between knowledge nodes, it would a solution to take them all into account.

9. References

[17] Vomlel