DYNAMIC LEARNING STYLE MODELLING USING PROBABILISTIC BAYESIAN NETWORK

Daiva Goštautaitė
Vilnius Gediminas Technical University (LITHUANIA)

Abstract

Personalised learning systems provide a unique, specific learning path for particular student or a group of students. They can adapt according to learner's requirements and preferences. They apply traditional information technologies, systems and tools in such a manner which provides learning based on student's strengths, weaknesses, psychological portrait, pace of learning, learner's needs and pedagogical methods best suited. Learning content personalisation, learning content type, representation of learning content, content navigation pattern are the main aspects to consider when personalising virtual learning environments. As personalisation is done by personal traits of a learner and by other information related to particular learner, user profiles and user models are used for modelling and storing such kind of information.

In this paper, first, a systematic review of literature on user modelling is done, focusing on static and dynamic user’s learning style models. Then Bayes approach to learning style modelling is introduced. In first subsection philosophical approach to representation of causality and belief is described – Bayes models are based on such approach. Then rules of probability theory applicable to Bayes models are presented. The following subsection is aimed at description of dynamic learning style modelling using probabilistic Bayes network.

Bayes network uses data about learner's past behaviour in web-based learning environment for prediction on properties to be used for future personalisation. As a lot of factors extracted from learner's past behaviour in adaptive hypermedia learning systems determine learning style [61], review of literature about patterns of learners' behaviour together with analysis of practical application of behavioural patterns for students learning style identification was done, trying to systematize stereotypical features (patterns) of learners’ behaviour that can be used to conclude a learning style. A list of key factors which probabilistically are related to the particular learning style has been compiled for quick handy use. Simulation of relationships between random key factors for learning style identification using Bayes probabilistic graphical model is also presented in the paper. Advantages and disadvantages of Bayesian learning style modelling were specified. Finally, conclusions and future trend are presented.

Keywords: virtual learning environments, personalisation, dynamic learning styles modelling, Bayes network, learners’ past behaviour.

1 INTRODUCTION

Learning in virtual learning environments, including web-learning systems, is everyday life in education. “One-size-fits-all” approach which is characteristic to many web-learning systems makes web-learning complex, and it is hard for students to orient themselves in such systems. Trying to solve this kind of problems, learning personalization was introduced. Personalised learning systems provide so called “user-centered” teaching by delivering content specific to the particular learner, presenting the content in an appropriate sequence, selecting type of content, providing guidance through the individual learning path, managing personalised views, providing recommendations of links, etc.

Adaptive hypermedia learning systems and intelligent tutoring systems are examples of techniques for providing personalised learning. Wide list of adaptive hypermedia learning systems was presented by Tomáš Kubeš in [8]. Most of today's adaptive systems are self-adjusting, i.e. their parameters depend on the history of the system dynamics. According to [7], we can expect regular behaviour of a system when we have complete information about its behaviour. In this case the Causality principle is valid. S. M. Korotaev in [43] explains, that principle of causality means retardation one event relatively another, i.e. necessary, but not sufficient condition of causal connection. To introduce sufficient condition means to define what observable is the cause and what is the effect. Effect of nonlocality compels to demarcate principle of strong causality and principle of weak causality. Principle of strong causality is: the cause precedes the all possible effects. Principle of weak causality is: the cause initiated by an
observer precedes the all possible effects. Besides, nonlocality violates principle of strong causality: the effects of the uncontrolled process-cause can forestall it [43].

The presence of randomness does not always imply violation of the principle. J.M. Rubi in [7] states, that a small amount of randomness not reaching chaos does not necessarily impede the formulation of causal laws in terms of probabilities. This is a case of so called weak causality. In [7] J.M. Rubi argue that despite of that the probabilities are involved, the law is also causal, and this causality refers to situations in which we have only a partial knowledge of the system. In [7] J.M. Rubi concluded that causality does not necessarily imply predictability but regularity, even for a probabilistic behaviour.

Such philosophical approach is applied in Bayes networks that are used as causal models of the real world. Wolfgang Spohn in [6] recalls German word for reality “Wirklichkeit”, which suggests, for something to be real it must stand in some causal relations, even if these are provided only by effects and not by causes. Bayesian networks are very convenient for representing probabilistic relationships between multiple variables (events, features, random processes), when the conditional independences between them can be stated. In other words, Bayesian networks represent alternatives of a hypothesis (states of a variable). They deal with uncertainty and are used for making future predictions and explaining observations.

Bayesian networks are called “belief” networks as they use probabilistic framework and update their “beliefs” when they receive new information as input. This is called belief propagation. Typically, belief change is determined by two factors: how reliable is new information and how well does the new information coheres with what it is already believed [9]. Brian Dennis in [11] treats Bayesian philosophy as scientific relativism, saying that truth is subjective, and there is no universal truth, only beliefs. In case of Bayesian network such beliefs are incorporated directly into the conclusion drawn from data [11].

Bayesian networks combine prior knowledge and data, manage situations where some data is missing or hidden and model causal relationships. After structural training, inferences about probabilities of different causes given the consequences might be made using Bayesian networks [21].

Bayesian networks could be used in adaptive hypermedia systems for user model prediction based on the historical user behaviour in the system. Recommender engine of personalised learning systems then can use preferences (for example, student's interests, learning style, etc.) inferred (predicted) by Bayesian network for recommendation of items a learner is interested (or preferred) most of all.

2 METHODOLOGY

As the aim of the paper was to make scientific research in the area of application of Bayesian methods for user modelling, systematic review of papers on the theme has been made. Literature based approach has been used for finding out learning personalisation issues and identification of student's behavioral factors determining learning style.

3 RESULTS

3.1 User modelling: current status, trends, possibilities

3.1.1 Personalisation concepts

In automated hypermedia learning systems personalization concepts of learner’s profile, learner’s stereotype and learner’s model are used. Learner’s profile could be defined as a collection of information which describes the learner, his/her traits, interests, preferences, competencies, his/her working environment, cultural features, etc. It consists of static user-specific data. Learner’s profile is associated with a single learner, and it uses information indicated by the learner himself/herself. Generally, stereotypes represent partial descriptions of frequently occurring situations. In this case, users are categorized according to stereotypical qualities. A complete model of a user can consist of various stereotypes which can intersect. Modelling via stereotypes means a generalized version of static profiles. Learner model’s purpose is to create widespread frame of reference for understanding a learner. Learner model can be continuously renewed, and it serves for automated learning system’s adaptation "on the go". Learner’s model typically uses information extracted from the data about student’s actions in adaptive hypermedia learning system by means of data mining. Learner’s model may be applied to many of learners.
3.1.2 Research related to learning style modelling

Paper [42] presents a survey on learner’s models in adaptive e-learning systems. Authors present explicit methods for generation of learner’s model: learning style modelling, standardization of learner’s information and use of ontologies. Comparative study between the most known learning style’s models is presented, giving the examples of e-learning systems that implement them. Further, authors analyse the most prominent standards adopted by computer-assisted learning systems (PAPI, IMS LIP, IMS RDECEO) and compare between them for learners’ information. Comparative study between the most recent ontologies for representing learner’s individual differences also is made and presented. Lastly, survey of implicit methods of learner’s modelling is presented. Authors state the main implicit method which is to analyse learner’s behaviour in virtual learning environment. After investigating navigational behaviour indicators, authors conclude that approaches explored didn’t consider navigational behaviour sufficiently. Some behavioural indicators are mentioned among those which have not been dealt with in the literature: educational preferences (habits, perception, etc.), indicators related to information processing (interactivity, difficulties understanding process, etc.) and cognitive abilities (motivation, rapidity, etc.). The conclusion is made that implicit deduction of learner’s model seems to be more satisfying than explicit. Insights presented in [29] could be used for creation of various learner models.

Authors of [32] describe user profiling trends, techniques and applications. They mention three types of user profiling: explicit, implicit and hybrid. After stating that, generally, user profiling system consists of profile extraction, profile integration and user interests (or characteristics) discovery, authors describe techniques for performing each task from the set of three. As profile extraction is an extraction of useful information about learner from log files and other sources, authors listed some learner’s activities that may be used for identification of learner’s interests. Text tracing, link clicking, link pointing, text selection, scrolling, bookmark registration, saving html document, printing a page, window movement or resizing are between them. According to [32], profile integration mainly refers to cleaning extracted data. In personalization context, learning content may be recommended to student according to his/her interests. For discovery of interests, content based filtering, collaborative filtering is applied.

It is known that one way to determine student’s learning style is by submitting a questionnaire to students and then using expert knowledge about learning styles for evaluation of answers to the questionnaire and classification of students. Authors in [41] state that the shortage in this case is that classification of students can be rough, not taking into account such aspects as mistakes in students’ answers, poorly prepared questionnaires, issues related to cultural environment, etc. The authors of [41] propose to solve these issues using Bayesian network for student classification into predefined classes according to learning style. Bayesian network would use historical data about student’s answers to questions in the questionnaire, thus making classification more accurate than one based on answers of a single student. For demonstration an example using Kolb’s learning cycle is presented [41].

In [25] predictive statistical models for user modelling are reviewed: linear models, Term Frequency Inverse Document frequency, Markov models, neural networks, classification, rule induction, Bayesian networks. It is stated that predictive performance using Bayesian network is better than using linear models with different weighting schemes or using Bayes classifier for a wide range of conditions.

In [24] Adaptive Bayesian network is proposed as predictive model for classification of learning resources as “appropriate” or “not appropriate” according to student’s learning style. Adaptive Bayes is an incremental learning algorithm that can work online. It is based on Naïve Bayes and includes updating schema, making possible to adapt to new data. Adaptive Bayes improves the probability associated with prediction due to the use of batch classifier (collect training data, then build a batch model with these data; drop an old model when memory is full; then repeat). The same adaptation strategy is used in Iterative Bayes. Two versions of adaptive Bayes exist: incremental and online. Authors evaluated both versions and concluded that use of Adaptive Bayes determines gains in prediction accuracy. Authors concluded that use of Adaptive Bayes is also a good choice in concept drift scenarios.

In [30] authors indicate that they decided to develop learning management system called “Suricata platform” which optimizes learning by personalizing learning process according to learner’s profile. It was decided to design a tool which proposes learning object for learner according to his/her “learning attitude”. According to [30], determination of learner’s profile is based on Bayes network. The model helps to determine in which styles a learner should be supported during his/her learning process.
In [38] authors proposed approach using Bayesian network for modelling student’s behaviour and inferring learning style according to it. In Bayesian network random variables are used for representation of factors that determine a particular dimension of learning style. Factors are extracted from the learner’s interactions with the system. Another authors in [2], [3] presented unified learning style model providing student’s behavioural indicators which could be used as factors influencing learning style. As Bayesian network is statistical graphical representation of learner model, simple probability tables for independent nodes (learning style dimensions) were set, initially assigning equal values and then adjusting to represent the new observations till the point when the probability values reach equilibrium. Conditional probability tables were set by combining expert knowledge and experimental results. When new evidence of student’s action is observed, values in probability tables are recalculated, using historical data about student’s actions and evidential data. Values of the nodes of Bayesian network corresponding to learning style’s dimensions given evidences of student’s behaviour are inferred, obtaining the marginal probability values of the learning style node given the values of independent nodes. Dominating student’s learning style is the one having the greatest probability values [38]. Experimental results are also presented and evaluated comparing them with the results obtained by means of filling by students the Felder and Silverman questionnaire. Precision for detecting student’s learning style is identified using similarity function. Authors concluded that Bayesian network enables to detect learning style with high precision.

Later authors of [38] improved their Bayesian model and described the enhancement in [27]. An enhancement concerns quantitative aspect of the model: probability function that corresponds to “processing” dimension of Felder and Silverman learning styles’ model was reformulated the way that encounters the influence of new factors. In other words, authors took into account the influence of different factors analysed on the dimensions of the learning style. Information about actual students’ behaviour during learning courses in the web was recorded and used together with expert knowledge for determining conditional parameters of the Bayesian network.

In [33] Bayesian network for learner’s preferences prediction is described. It maintains knowledge about the interactions, habits, preferences of the learner. Preferences then are used for adaptation purposes as they determine which applications are appropriate to the user, which documents to show, which graphical components should be included in the user interface, etc. Simulated annealing algorithm has been used to train the structure of the network: current solution is moved to less optimal states based on a probability function, preventing a local optimum from restricting the algorithm. For computing posterior probability distributions for all variables junction tree algorithm has been used.

In [39] authors present the combination of an overlay model used for learner’s mastery of the particular knowledge representation with Bayesian network used for probabilistic reasoning. Overlay model is a subset of domain/knowledge model. In spite of that authors stressed that hypothesis to combine overlay model and Bayesian network is appropriate approach to learner modelling, they indicated that this approach may be difficult if a high amount of data is stored system. To overcome this problem multi-entity Bayesian networks combining Bayesian network with first-order logic are proposed.

Another combination of Bayesian network and overlay model is described in [28]. It is used for student’s knowledge diagnosis. Here the structure of the overlay model is considered as Bayesian network. All student’s knowledge elements in the network are defined as hidden variables (hypothesis). Learning objects or events which are used to evaluate user’s performance in learning process are considered as evidence variables. Presumptions that variables are mutually exclusive, random and binary are made according to Bayesian philosophy. Development of qualitative model is done by experts and teachers. Development of quantitative model (specifying conditional probability tables) is made using weights of prerequisite relationships. If concept has no prerequisite knowledge for understanding, it’s values of conditional probabilities in the table are specified as prior probabilities obeying uniform distribution (assigned values 0.5). Inferences about learner’s knowledge of concept are made according to calculated posterior probability conditional of hidden variables. Having no doubts about appropriateness of the approach for user modelling, authors also highlight two main disadvantages of it: the expense of data storage in case number of parameters is high and computational time.

Loc Nguyen and Bich-Thuy T. Dong in [40] presented triangular model of learner’s characteristics used by user modelling system ZEBRA. This model consists of three sub-models: knowledge sub-model, learning history sub-model and learning style sub-model. Learning style of student is modelled using Hidden Markov model, and it is considered as a state. State transition is a change in student’s learning style. Learning style is discovered using observations and inference mechanism of Hidden
Markov model. The hidden Markov model can be represented as the simplest dynamic Bayesian network. According to [40], the purpose of ZEBRA is to mine user’s learning profile, so it has mining engine. User modelling is similar to profile mining. Model has belief network engine able to reason according to evidence. Thus, belief network is responsible for inferring new student’s knowledge and learning styles from triangular learning model by using deduction mechanism in belief network. This engine applies both Hidden Markov model and Bayesian network. Content selection in the system is being made by matching learning style with attributes assigned to content resources. There are two daemons using Bayesian networks in the system. Note, that daemon is a program that runs continuously, and it exist for a purpose of handling periodic service requests that system expects to receive. It forwards the requests to other programs or processes as appropriate. Knowledge static daemon evaluates student’s knowledge based on Bayesian mechanism. Knowledge dynamic daemon models student’s knowledge in chronological process. There are also daemons related to learning history (log files) management, learning history data mining for extraction of patterns and breaking them into learning material which is recommended to the learner. Triangular learning model and the system proposed in [40] both support ubiquitous user modelling. It enables ongoing modelling, ongoing sharing and ongoing exploitation of user models in ubiquitous environment that shifts from desktop interaction to mobile interaction. Ubiquitous user modelling is considered as intersection of three domains: user modelling, ubiquitous computing and semantic web. Ubiquitous user modelling system simulates the real world as virtual world. It is defined as the user model which is monitored at any time, at any location and in any interaction context. Ubiquitous user model can be shared or integrated when necessary [40].

In [42] authors apply literature-based approach for learning style detection from learner’s behaviour in web based learning environment: after patterns have been defined for particular dimension of learning style and threshold levels assigned to the patterns, they are used as evaluation criteria for learner’s behaviour data, thus detecting student’s learning style automatically.

In [1] a proposal to use tree augmented naïve Bayes network for learning style detection is made. Augmented Naive Bayes network maintains the robustness and computational complexity of the Naive Bayes model and at the same time displays better accuracy. Computational complexity of this model is significantly reduced as each variable in Augmented Naive Bayes network has a maximum of two parents. Beside of that, Augmented Naive Bayes network is more realistic as it loosens the strong independence assumptions used in Naive Bayes networks, at the same time imposing restrictions on the level of interaction between the variables.

Regarding experimentation in learning style modelling using Bayesian approach, it should be stated that experiments described in the papers are only small scale trials, frequently consisting of one to tree learning style’s dimensions and only few behavioural activities influencing them.

In [26] authors analysed articles on adaptation according to learning style published between 2001 and 2016. Authors stated positive perspectives for correct automated prediction of learning styles.

### 3.1.3 Factors determining learning style

Literature [1]-[48] on data mining and learning style modelling was reviewed trying to identify factors that influence determination of student’s learning style. The actual values of these factors could be deduced from the students’ interaction with virtual learning environment by means of tracking student’s behaviour in the environment. Comparative analysis was made trying to compare and systematize factors described in the literature. Most of the papers reviewed deal with factors which determine learning styles of only one learning style model chosen, except the [2] paper, authors of which propose unified learning model (ULSM) approach. As adaptive hypermedia systems differ in technologies and devices available and standards used, besides, learners differ in the content of courses they are studying, it can be concluded that both ULSM and whichever single model are relevant and applicable depending on the concrete case. With such point of view, the purpose was to present a framework of systematised stereotypical features (or/and patterns) of learner’s behaviour (called factors) that determine learning style. These factors may be used in student’s learning style simulation resulting in dynamic learning style model. The aim of the framework is to present for quick handy use a list of key factors which probabilistically are related to the particular learning style. As full list is of 5 pages in length only short passage of the framework is presented in Table 1 – full list of factors can be obtained from the author.
### Table 1. Passage of the list of behavioural factors determining learning style.

<table>
<thead>
<tr>
<th><strong>PERCEPTION: sensors</strong></th>
<th><strong>PERCEPTION: intuitive students</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dealing with concrete content</strong></td>
<td><strong>Dealing with abstract content</strong></td>
</tr>
<tr>
<td>adding links that contains facts</td>
<td>make lists of key concepts from material</td>
</tr>
<tr>
<td>access of examples (concrete content) first</td>
<td>intuitive leaps through content: fast clicks on course material</td>
</tr>
<tr>
<td>higher number of exercises, practical tests done</td>
<td>frequent looking for hints</td>
</tr>
<tr>
<td><strong>Notices details</strong></td>
<td><strong>Deep, imaginative, theoretical, inventive</strong></td>
</tr>
<tr>
<td>repeated exercises revision, higher number of exam revisions in relation to the time of exam</td>
<td>more frequent multitasking: higher number of tabs opened in browser, frequent transition from one browser to the other</td>
</tr>
<tr>
<td>much exam delivery time</td>
<td>frequent use of conceptual maps</td>
</tr>
<tr>
<td>much time spent on viewing the exam sample files first time</td>
<td>frequent transition from one subject to the other</td>
</tr>
<tr>
<td><strong>Goes by senses, lives in the present</strong></td>
<td>frequent use of quiz tests which can be automatically marked</td>
</tr>
<tr>
<td>choose of tests with multiple-choice questions</td>
<td>higher number of tests retaken</td>
</tr>
</tbody>
</table>

For representation of relationships between random factors and students’ learning style dimensions, probabilistic graphical model – Bayesian network – may be used. Bayesian network facilitates efficient inferences about learning style from learners’ behavioural observations [5]. Author of [4] emphasizes that Bayesian network represents relations of probability, not necessarily causality, and what distinguishes a causal relationship from a merely probabilistic one is that it supports intervention. It must be stressed, that inferences about student’s learning style in Bayesian network are made in an evidential (not directly causal) mode of reasoning.

### 3.2 Bayesian model for student’s learning style modelling

#### 3.2.1 Probabilistic model used in Bayesian modelling

A Bayesian network is a directed acyclic graph in which edges correspond to conditional dependencies (probabilistic dependencies between the nodes they connect) and nodes correspond to unique random variables. For each edge there is conditional probability table defined for a set of discrete and mutually exclusive random variables to display conditional probabilities of a single variable with respect to the others. Conditional probability table grows exponentially with number of parents. To avoid this exponential growth, the inference process in a Bayesian network uses local distributions of one random variable depending on its parents. As this reduces the number of parameters used in calculation of probability, statistical inferences become tractable [21].

Bayesian networks represent the distribution of multiple, discrete or continuous, random variables [21]. Network nodes send probabilistic information according to the rules of probability theory (Bayes’ theorem). Bayesian network’s global semantics defines the full joint distribution as the product of the local conditional distributions. Calculation of any member of the joint distribution of a set of random variables is made using conditional probabilities according to chain rule. A Bayesian network’s local semantics means that each node is conditionally independent of its non-descendants given its parents.

In Bayesian networks we can exchange complexity with accuracy: making dependencies week in order to restrict complexity results in a lower accuracy.

For learning style identification using Bayesian network, simulation of relationships between random key factors should be modelled. In general, more than one factor (random variable \(x_i\)) influences each dimension of learning style. Therefore, probability distribution \(P(X_1, X_2, \ldots, X_n)\) should be calculated, using chain rule [21]:

\[
P(X_1, X_2, \ldots, X_n) = P(X_1) \prod_{k=2}^{n} P(X_k | X_{k-1}, \ldots, X_1)
\]  

(1)

Conditional independency can be used to rewrite the chain rule to:
\[ P(X_1, X_2, \ldots, X_n) = P(X_1) \prod_{i=2}^{n} P(X_i | P(X_i)) \]  

(2)

Here \( P(X_i) \subseteq \{X_1, X_2, \ldots, X_{i-1}\} \) are called the parents of \( X_i \). Variables which are not in the set of parents are assumed to be conditionally independent of \( X_i \). When a new evidence is observed, the probability distribution is recalculated [21].

Rainer Deventer [21] states that Bayesian network might be used for two main purposes. The first one is called marginalization [21]. Marginalization requires summing over the possible values of one variable to determine the marginal contribution of another. It tells us that we can calculate the quantity we want if we sum over all possibilities [22]. Note that in case of dynamic Bayesian network with continuous random variables summation is replaced by integration. The second purpose is for a case of new evidence, and it answers the question how is the probability distribution \( P(X) \) recalculated, when new evidence is observed, i.e. how to calculate \( P(X|X_i = x_i) \), where \( X_i = x_i \) is new evidence [21].

In summary, for the Bayesian network to make inferences, it is necessary to calculate marginal distributions, include evidence, or to calculate instantiations of the random variables which lead to a maximal probability [21]. For the propagation of evidence Junction tree algorithm is used in most cases [21]. Hard, virtual and soft evidences are distinguished in Bayesian network.

3.2.2 Bayesian network construction and training

Two basic approaches are used to construct Bayesian networks [20]: data-based and knowledge-based approaches. Data-based method uses conditional independence semantics of Bayes networks to make inferences from data whereas the knowledge-based approach utilizes causal knowledge from domain experts to construct Bayesian network.

Before training Bayesian network only data about the domain being modelled are given, i.e. neither the structure, nor the conditional probabilities are known. According to [21], when learning the parameters of a distribution in the Bayesian network, it is of advantage, if all nodes are observed. Besides, the usage of hidden nodes, which do not represent an existing value, is sometimes helpful in order to reduce the number of parameters. The data of the modelled processes also can contain continuous variables. Thus, it is necessary to use a training method which is able to deal also with continuous values, ideally with discrete and continuous values at the same time [21].

In Bayesian network uncertainty is assessed in point estimates (posterior distribution’s statistics of central tendency). It is desirable for a point estimate to be consistent (the larger the sample size, the more accurate the estimate) and unbiased (possessing the smallest variance) [36]. Few approaches are applicable for Bayesian network learning (calculation of the estimator): maximum likelihood estimation, moment’s method, Bayesian approach [36]. Expectation maximization algorithm is frequently used for training with missing data. It is based on maximum likelihood estimation and is able to deal with discrete and continuous data at the same time [21].

3.2.3 Learning style modelling using Bayes network

Using Bayesian networks, any question that can be posed in a probabilistic form can be answered with certain level of confidence [34]. Problems of classification, segmentation, state estimation, fault diagnosis, prediction might be solved using Bayesian model. Bayesian networks might be used to find the probability of independent variable being in a certain state, given certain values of other variables in the network or to find values of a particular set of variables that best explains why set of other variables have certain values [34]. Student’s learning style modelling belongs to the first task. Thus, to predict student’s learning style, it is needed:

- to build and learn generative model from historical data about student’s behavioural activities, influencing respective learning style dimensions,
- using Bayes rule, compute the probability that the particular learning style dimension is characteristic to the student, given new evidence of student’s behaviour.

Pilot experimental learning style simulation was done using GeNie Academic Version 2.3.3828.0. Model is presented in Figure 1. Knowledge based approach for network training was applied.
3.2.4 Advantages and disadvantages of Bayesian application to data modelling

Before deciding to apply Bayes network for a particular task, Bayes network’s advantages and drawbacks should be considered. In [20] Pekka Kekolahti and other authors among parametric Bayesian modelling advantages stated possibilities to model sequential data and causal relationships, to combine prior knowledge and data [20], to use Bayes’ conditionalization mechanism for updating belief in light of new evidence, to manage situations where some data is missing [20], possibility of structural learning of Bayes network, possibility of making inferences about probabilities of different causes given the consequences [20], possibility to incorporate different types of information.

The following general disadvantages of parametric Bayesian modelling were identified:

- difficulty to obtain prior knowledge in a form that can be converted into probability distributions,
- Bayesian network only encodes directional relationship (not bi-directional) – cyclic relationship is not allowed; there is a lack of support of feedback loops due to acyclic nature of network [20],
- Bayesian network reflects impact, but not the cause and effect relationship,
- typically, data to be modelled should be independent and identically distributed,
- inferences are made using past historical data only – Bayesian network doesn’t know how to react to unforeseen situations; here so called “cold start” problem should be mentioned,
- inferences are subjective and probabilistic, they are useful only in cases when prior knowledge is reliable and are not suitable for science where strong causality cases are under consideration,
- inferences are brute: in reality, there is a low likelihood of independence of variables,
- joint probability distribution may be biased when it is not modelling complete data, i.e. it may only be suitable under closed-world assumptions [45],
- to make relationships between variables to reflect reality, huge load is necessary,
- typical Bayesian network is finite dimensional model having finite number of parameters, therefore it is not flexible; for some problems nonparametric Bayesian models are more suitable,
- continuous variables are handled only in a limited manner [20], using Dynamic Bayes networks,
• there is a high computational cost in models with a large number of parameters; learning such a Bayesian network (selecting a probabilistic model that explains a given set of data [34]) is complex. There is no efficient deterministic way to find the best network, and it might be needed randomized algorithms like Markov chain Monte Carlo to find a good network. Bayesian network's structure learning is NP-complete problem,

• simulations provide slightly different answers unless the same random seed is used [35],

• simple Bayes network models only linear hypothesis and likelihood distributions - nonlinear factor analysis and nonlinear state-space models are only special cases of Bayes networks where the linearity assumption is relaxed,

• to use a probability distribution function alone to represent uncertainty is not enough, because it fails to show the ignorance (called “sensitivity”), or uncertainty about the function itself, i.e. the Bayesian approach has no general way to represent and handle the uncertainty within the background knowledge and the prior probability function [37]. Pei Wang in [37] emphasizes that if we want to apply a Bayesian network to a practical domain, one of the following requirements must be satisfied:

• the implicit condition of the initial probability distribution, that is, the domain knowledge used to determine the distribution initially, can be assumed to be immune from future modifications, or

• all modifications of the implicit condition can be treated as updating, in the sense that when new knowledge conflict with old knowledge, the latter is completely abandoned.

Bayesian network is not able to combine conflicting beliefs that are based on different implicit conditions and to carry out inference when premises are based on different implicit conditions. Confusion between explicit and implicit conditions of probability evaluations (as well distinction between revision and updating) causes underestimation about the limitation of “Bayesianism” [37].

4 CONCLUSIONS

Bearing in mind the overall strengths and weaknesses of the Bayesian network, we draw a conclusion that Bayesian network is a right choice for student's learning style modelling. As learning style-setting student's behavioural activities in hypermedia learning environments represent mostly discrete and independent streaming data, simple parametric Bayesian network using discrete distributions that represent belief states may be usable.

As stated in [46], Naive Bayes modelling makes an assumption that all the attributes (features) are conditionally independent. For real cases when it is necessary to model directly or indirectly dependent of each other attributes, an extension of the Naive Bayes model that also includes some (conditional) dependencies between the random variables might be usable [47]. An example is augmented Naive Bayes network which augments network by adding edges between the attributes to include the information of interdependence between the attributes [47]. To reduce computational complexity in this case, restriction on the level of interaction between variables can be imposed. Tree-Augmented Naive Bayes Model in which each variable has only one or two parents, except for the root variable, belongs to the category of such models.

More accurate than simple Bayes models should learn as the new sample data arrives – adaptive Bayesian networks may be used for that purpose [24], [45], [46]. This category of Bayesian models can handle concept drift: detect, when concept change has occurred, and adapt to the changes accordingly [24]. According to [24], adaptive Bayesian networks also handle the trade-off between cost of updating and the gain in performance, deciding whether it is necessary to initiate updating process.

REFERENCES


