

A BAYESIAN METANETWORK

VAGAN TERZIYAN

*Department of Mathematical Information Technology, University of Jyväskylä,
P.O. Box 35 (Agora), FIN-40014 Jyväskylä, Finland
vagan@it.jyu.fi*

Received (9 July 2003)
Accepted (15 March 2004)

Bayesian network (BN) is known to be one of the most solid probabilistic modeling tools. The theory of BN provides already several useful modifications of a classical network. Among those there are context-enabled networks such as multilevel networks or recursive multinets, which can provide separate BN modelling for different combinations of contextual features' values. The main challenge of this paper is the multilevel probabilistic meta-model (Bayesian Metanetwork), which is an extension of traditional BN and modification of recursive multinets. It assumes that interoperability between component networks can be modeled by another BN. Bayesian Metanetwork is a set of BN, which are put on each other in such a way that conditional or unconditional probability distributions associated with nodes of every previous probabilistic network depend on probability distributions associated with nodes of the next network. We assume parameters (probability distributions) of a BN as random variables and allow conditional dependencies between these probabilities. Several cases of two-level Bayesian Metanetworks were presented, which consist on interrelated predictive and contextual BN models.

Keywords: Bayesian networks, context, multinets

1 Introduction

A *Bayesian network* (BN) has proven to be a valuable tool for encoding, learning and reasoning about probabilistic (causal) relationships [1]. A BN for a set of variables $X = \{X_1, \dots, X_n\}$ is a directed acyclic graph with a network structure S that encodes a set of conditional independence assertions about variables in X , and a set P of local probability distributions associated with each variable [2]. Simple BN example is shown in Fig. 1.

Inference in BN generally targets the calculation of some probability of interest. Inference algorithms exploit the conditional (in)dependence between variables (1), see e.g. joint probability (2) for Fig. 1; marginalization rule (3); and Bayesian rule (4).

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Parents}(X_i)) \quad (1)$$

$$P(Y = y_j, X = x_i) = P(X = x_i) \cdot P(Y = y_j | X = x_i) \quad (2)$$

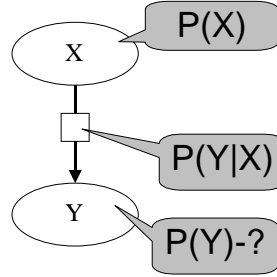


Fig. 1. Example of a simple Bayesian network

$$P(Y = y_j) = \sum_i P(X = x_i) \cdot P(Y = y_j | X = x_i) \quad (3)$$

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i) \cdot P(Y = y_j | X = x_i)}{P(Y = y_j)} \quad (4)$$

Learning BN generally means the refinement of the structure and local probability distributions of a BN given data. The simplest version of this problem is using data to update the probabilities of a given BN network structure. An important task in learning BN from data is model selection [3]. Each attribute in ordinary BN has the same status, so they are just combined into possible models-candidates to encode conditional dependencies. Some modifications of BN however require distinguishing between attributes, e.g. as follows:

- *Target attribute*, which probability is being estimated based on set of evidence.
- *Predictive attribute*, which values being observed effect probability distribution of a target attribute(s) via some structure of other predictive attributes according to causal dependencies among them.
- *Contextual attribute*, which has no direct visible effect to target attributes but influences some of probability distributions within the predictive model. A contextual attribute can be conditionally dependent on some other contextual attribute.

Causal independence in a BN refers to the situation where multiple causes provided by predictive attributes contribute independently to a common effect on a target attribute. With causal independence, the probability function can be described using a binary operator that can be applied to values from each of the parent predictive attributes (see (1)). Context specific independence refers to the fact that some random variables are probabilistically independent of each other in a certain context.

In [4], Butz exploited contextual independencies based on assumption that while a conditional independence must hold over all contexts, a contextual independence need only hold for one particular context. He shows how contextual independencies can be modeled using multiple BN. Boutilier et al [5] presents two algorithms to exploit context specific independence in a BN. The first is network transformation and clustering. With this method, the context specific independencies are qualitatively encoded within the

network structure of the transformed network and appropriate conditional probability tables are represented as decision trees. The other algorithm works by selecting a set of variables that, once instantiated, makes the network singly connected. With context specific independence this essentially reduces total inference time for queries. Zhang [6] used generalized rules for contextual variable elimination algorithm, which capture context specific independence in variables. Geiger and Heckerman [7] used similarity networks to make context specific independencies explicit in BN. In a similarity network, the edges between the variables can be thought of as a network whose edges can appear or disappear depending on the values of certain variables in the network. This allows for different BNs to perform inference for different contexts.

Domain ontologies seem to be an excellent source of contextual data, which can be used for modelling context-sensitive BNs. Additional knowledge about (semantic) relations between BN nodes other than causal ones can be considered as a sample of a very useful context, which can affect the interpretation of the original structure of the BN. Helsen and Gaag [8] describe an interesting combination of medical ontologies as explicit documentation of the cancer domain knowledge and the BN, which encodes cause-symptom relationships in this domain. For example, the knowledge that pertains to *haematogenous metastases* may be considered from different points of view and resulting in the two alternative depictions (see Fig. 2).

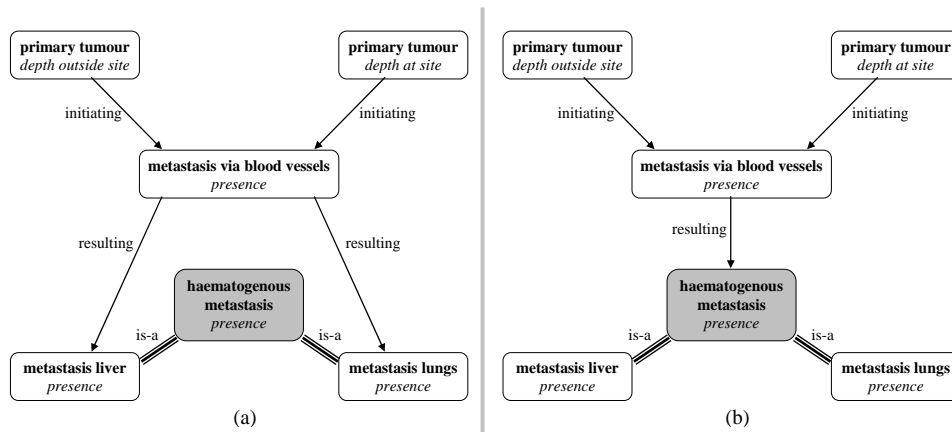


Fig 2. Combining BN with ontology [9]

Alternative (a) in Fig. 2 describes that the process of *metastasis via blood vessels* may result in *metastases in the lungs* and *metastases in the liver*, which are known from the ontological context to be subclasses of the class *haematogenous metastasis*. Alternative (b) captures a relation at a higher level: the process of *metastasis via blood vessels* may result in *haematogenous metastases*, which according to the context may be in the lungs or in the liver. Imagine a scenario when some BN learning algorithm has produced structure (a). Imagine also that our context as part of OWL ontology is knowledge about: class *haematogenous metastasis* is a disjoint_union_of class *metastases in the lungs* and class *metastases in the liver*. According to this context it is possible to automatically transfer BN structure from alternative (a) to more compact alternative (b) and then make appropriate recalculation of the BN parameters.

If a Bayesian network is to be modified to better represent probability of a target attribute, one can either change its graphical structure, or its parameters or both [9]. There are two ways for changing the structure of a Bayesian network: one can change the nodes in the graph (add, delete or combine nodes) or change the arrows in the graph (add, delete or re-orient arrows) or both. In Bang et al. [9] a combination of these two strategies for network modification is considered. To generate a causal network that satisfies the causal Markov condition the flexibility is needed to add either a new arrow or a new common cause (a hidden node). The complexity of generated network can be measured in terms of the number of parameters required in the network. Results show that in most situations adding an arrow will increase complexity least, but in cases where two or more nodes share a large number of parents, adding a hidden node can even decrease complexity.

In [10] a multi-level BN was presented that accurately models the system and allows for sensor integration in an evidential framework. It was shown that a multi-level BN performs better than a simple single-level BN. Multilayer networks are usually sensitive to conditional probabilities, which should be defined with greater accuracy because small differences in their values may result in radically different target concept estimation. Choosing between a simple BN and a multilevel network one needs to carefully evaluate an expected benefit against the increased costs of the knowledge management [11].

Bayesian multinets were first introduced in [12] and then studied in [13] as a type of classifiers. A Bayesian multinet is composed of the prior probability distribution of the class node and a *set* of local networks, each corresponding to a value that the class node can take. Bayesian multinets can be viewed as a generalization of BNs. A BN forces the relations among the features to be the same for all the values that the class node takes; by contrast a Bayesian multinet allows the relations among the features to be different, i.e. for different values the class node takes, the features can form different local networks with different structures. While multinet is more general than BN, it is often less complex since some of the local networks can be simpler than others, while BN needs to have a complex structure in order to express all the relationships among the features [13]. In [14] dynamic Bayesian multinets are introduced where a Markov chain state at time determines conditional independence patterns between random variables lying within a local time window surrounding it. It is demonstrated that multinets can in many cases outperform other dynamic models with a similar number of parameters. A recursive Bayesian multinet was introduced by Pena et al [15] as a decision tree with component Bayesian networks at the leaves. The key idea was to decompose the learning Bayesian network problem into learning component networks from incomplete data.

As our main goal in this paper, we are presenting another view to the Bayesian “multinets” towards making them to be really “metanetworks”, i.e. by assuming that interoperability between component Bayesian networks can be also modeled by another Bayesian network.

The rest of paper organized as follows. In Section 2 we present two models of a contextual effect to probability distributions in BN. In Section 3 we introduce a Bayesian Metanetwork, basic inference in it and few cases of interaction between its predictive and contextual levels. We conclude in Section 4.

2 Modelling Contextual Effect on Bayesian Network Parameters

In this chapter, we first consider the model of direct effect of a contextual parameter on conditional (2.1) and unconditional (2.2) probabilities in a BN (or actually indirect effect to a target attribute's probability), which is one of the basic components of the Bayesian Metanetwork concept.

2.1. Contextual Effect on Conditional Probability

Simplest possible case is shown in Fig. 3 where one contextual attribute affects the conditional probability between one predictive and one target attribute.

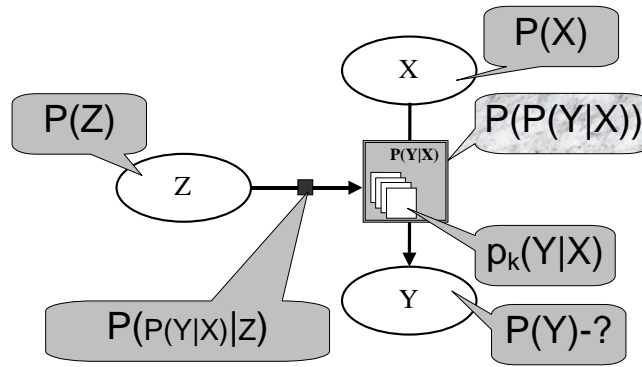


Fig. 3. Simple model of contextual effect on conditional probability

The case shown in Fig. 3 can be described as follows:

- $X = \{x_1, x_2, \dots, x_n\}$ – predictive attribute with n values;
- $Z = \{z_1, z_2, \dots, z_q\}$ – contextual attribute with q values;
- $P(Y|X) = \{p_1(Y|X), p_2(Y|X), \dots, p_r(Y|X)\}$ – conditional dependence attribute (random variable) between X and Y with r possible values;
- $P(P(Y|X)|Z)$ – conditional dependence between attribute Z and attribute $P(Y|X)$;

Assume that our goal is to compute $P(Y)$. For that first we calculate the probability for the “conditional dependence” attribute:

$$P(P(Y | X) = p_k(Y | X)) = \sum_{m=1}^q \{P(Z = z_m) \cdot P[P(Y | X) = p_k(Y | X) | Z = z_m]\} \quad (5)$$

Then we estimate the following joint probability:

$$\begin{aligned} &P(Y = y_j, X = x_i, P(Y | X) = p_k(Y | X)) = \\ &= P[Y = y_j | X = x_i, P(Y | X) = p_k(Y | X)] \cdot P[P(Y | X) = p_k(Y | X)] \cdot P(X = x_i) \end{aligned} \quad (6)$$

Taking into account that:

$$P[Y = y_j | X = x_i, P(Y | X) = p_k(Y | X)] = p_k(Y = y_j | X = x_i)$$

we can rewrite (6) as follows:

$$\begin{aligned}
& P(Y = y_j, X = x_i, P(Y | X) = p_k(Y | X)) = \\
& = p_k(Y = y_j | X = x_i) \cdot P[P(Y | X) = p_k(Y | X)] \cdot P(X = x_i)
\end{aligned} \tag{7}$$

Substituting (5) to (7) we get:

$$\begin{aligned}
& P(Y = y_j, X = x_i, P(Y | X) = p_k(Y | X)) = \\
& = p_k(Y = y_j | X = x_i) \cdot P(X = x_i) \cdot \sum_{m=1}^q \{P(Z = z_m) \cdot P[P(Y | X) = p_k(Y | X) | Z = z_m]\}
\end{aligned} \tag{8}$$

Applying marginalization in (8) we obtain:

$$\begin{aligned}
& P(Y = y_j) = \sum_{k=1}^r \sum_{i=1}^n \{p_k(Y = y_j | X = x_i) \cdot P(X = x_i)\} \times \\
& \times \sum_{m=1}^q [P(Z = z_m) \cdot P(P(Y | X) = p_k(Y | X) | Z = z_m)]
\end{aligned} \tag{9}$$

or in more compact form (with attributes only and without values) the general calculation scheme will be as follows:

$$P(Y) = \sum_{\forall P(Y|X)} \sum_{\forall X} \{P(X) \cdot P(Y | X) \cdot \underbrace{\sum_{\forall Z} [P(Z) \cdot P(P(Y | X) | Z)]}_{\text{metalevel}}\}. \tag{10}$$

Consider artificial example scenario for Fig. 3:

- *Target attribute (someone's wellness):* $Y = \{\text{Rich, Poor}\}$;
- *Predictive attribute (someone's intension to work):* $X = \{\text{Hardworking, Lazy}\}$;
- *Known probability distribution:* $P(\text{Hardworking}) = 0.3$; $P(\text{Lazy}) = 0.7$.
- *Contextual attribute (someone's country of residence):* $Z = \{\text{USA, Ukraine, TheRestWorld}\}$;
- *Probability distribution:* $P(\text{USA}) = 0.2$; $P(\text{Ukraine}) = 0.1$; $P(\text{TheRestWorld}) = 0.7$.
- Assume that we know two possible conditional probability distributions for $P(Y|X)$ as it is shown in Table 1.

Table 1. Probability distribution for $P(X|Y)$ in the example

$p_1(Y X)$	Hardworking	Lazy
Rich	0.8	0.1
Poor	0.2	0.9
$p_2(Y X)$	Hardworking	Lazy
Rich	0.6	0.5
Poor	0.4	0.5

- Assume conditional dependence between contextual attribute Z and $P(Y|X)$ to be as shown in Table 2.

Table 2. Conditional dependence between context (attribute Z) and conditional probability P(Y|X) in the example

P(P(Y X) Z)	USA	Ukraine	TheRestWorld
$p_1(Y X)$	0.9	0.2	0.7
$p_2(Y X)$	0.1	0.8	0.3

- So we have: $n = 2$; $q = 3$; $r = 2$.

Assume that task is to calculate P (Y), i.e. P(Y=Rich) and P(Y=Poor).

First calculate the following according to (5):

$$\sum_{m=1}^3 [P(Z = z_m) \cdot P(P(Y | X) = p_1(Y | X) | Z = z_m)] = 0.2 \times 0.9 + 0.1 \times 0.2 + 0.7 \times 0.7 = 0.69;$$

$$\text{Similarly: } \sum_{m=1}^3 [P(Z = z_m) \cdot P(P(Y | X) = p_2(Y | X) | Z = z_m)] = 0.31.$$

Now we apply (9):

$$P(Y = Rich) = 0.8 \times 0.3 \times 0.69 + 0.1 \times 0.7 \times 0.69 + 0.6 \times 0.3 \times 0.31 + 0.5 \times 0.7 \times 0.31 = 0.3782;$$

$$\text{Similarly: } P(Y = Poor) = 0.6218.$$

2.2. Contextual Effect on Unconditional Probability

Simplest possible case is shown in Fig. 4 where one contextual attribute affects unconditional probability distribution of a predictive attribute.

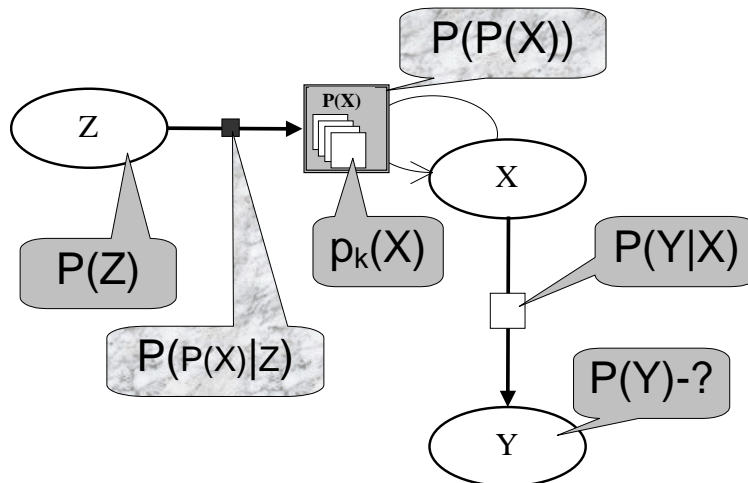


Fig. 4. Simple model of contextual effect on unconditional probability

In the case shown in the Fig. 4 we have the following input data:

- $X = \{x_1, x_2, \dots, x_n\}$ – predictive attribute with n values;
- $Z = \{z_1, z_2, \dots, z_q\}$ – contextual attribute with q values and $P(Z)$ – probability distribution for values of Z ;
- $P(X) = \{p_1(X), p_2(X), \dots, p_r(X)\}$ – probability distribution attribute for X (random variable) with r possible values (different possible probability distributions for X) and $P(P(X))$ is probability distribution for values of attribute $P(X)$;
- $P(Y|X)$ is a conditional probability distribution of Y given X ;
- $P(P(X)|Z)$ is a conditional probability distribution for attribute $P(X)$ given Z ;

Assuming that goal is to compute $P(Y)$, first we calculate the following probability:

$$P(P(X) = p_k(X)) = \sum_{m=1}^q \{P(Z = z_m) \cdot P[P(X) = p_k(X) | Z = z_m]\}. \quad (11)$$

Then we estimate the following joint probability:

$$\begin{aligned} P(Y = y_j, X = x_i, P(X) = p_k(X)) &= \\ &= P(Y = y_j | X = x_i, P(X) = p_k(X)) \cdot P(X = x_i, P(X) = p_k(X)) = \\ &= P(Y = y_j | X = x_i, P(X) = p_k(X)) \cdot P(X = x_i | P(X) = p_k(X)) \cdot P(P(X) = p_k(X)) \end{aligned} \quad (12)$$

Taking into account the independence of parameters Y and $P(X)$ given X and also that:

$$P(X = x_i, P(X) = p_k(X)) = p_k(X = x_i)$$

we can rewrite (12) as follows:

$$\begin{aligned} P(Y = y_j, X = x_i, P(X) = p_k(X)) &= \\ &= P(Y = y_j | X = x_i) \cdot P(P(X) = p_k(X)) \cdot p_k(X = x_i) \end{aligned} \quad (13)$$

Substituting (11) to (13) we get:

$$\begin{aligned} P(Y = y_j, X = x_i, P(X) = p_k(X)) &= \\ &= P(Y = y_j | X = x_i) \cdot p_k(X = x_i) \cdot \sum_{m=1}^q (P(Z = z_m) \cdot P(P(X) = p_k(X) | Z = z_m)) \end{aligned} \quad (14)$$

Applying summarization in (14) we obtain:

$$\begin{aligned} P(Y = y_j) &= \sum_{k=1}^r \sum_{i=1}^n \{P(Y = y_j | X = x_i) \cdot p_k(X = x_i) \times \\ &\times \sum_{m=1}^q [P(Z = z_m) \cdot P(P(X) = p_k(X) | Z = z_m)]\} \end{aligned} \quad (15)$$

or in more compact form the general calculation scheme will be as follows:

$$P(Y) = \sum_{\forall P(X)} \sum_{\forall X} \{P(X) \cdot P(Y | X) \cdot \underbrace{\sum_{\forall Z} [P(Z) \cdot P(P(X) | Z)]}_{\text{metalevel}}\}. \quad (16)$$

Consider artificial example scenario for Fig. 4:

- *Target attribute (someone's wellness):* $Y = \{\text{Rich, Poor}\}$;
- *Predictive attribute (someone's intension to work):* $X = \{\text{Hardworking, Lazy}\}$;
- Assume that we have two possible probability distributions for X as it is shown in Table 3.

Table 3. Two probability distributions for X in the example

P(X)	Hardworking	Lazy
$p_1(X)$	0.2	0.8
$p_2(X)$	0.5	0.5

- *Contextual attribute (someone's country of residence):* $Z = \{\text{USA, Ukraine, TheRestWorld}\}$;
- Probability distribution: $P(\text{USA}) = 0.2$; $P(\text{Ukraine}) = 0.1$; $P(\text{TheRestWorld}) = 0.7$.
- Assume that we know conditional probability distribution matrix for $P(Y|X)$ as it is presented in Table 4.

Table 4. Conditional probability $P(Y|X)$ in the example

P(Y X)	Hardworking	Lazy
Rich	0.7	0.4
Poor	0.3	0.6

- Let conditional dependence between contextual attribute Z and attribute P(X) to be as it is shown in Table 5.

Table 5. Conditional dependence between context (attribute Z) and probability distribution P(X) in the example

P(P(X) Z)	USA	Ukraine	TheRestWorld
$p_1(X)$	0.4	0.2	0.3
$p_2(X)$	0.6	0.8	0.7

- So we have: $n = 2$; $q = 3$; $r = 2$.

Assume that task is to calculate $P(Y)$, i.e. $P(Y=\text{Rich})$ and $P(Y=\text{Poor})$.

First we calculate the following according to (11):

$$\sum_{m=1}^3 [P(Z = z_m) \cdot P(P(X) = p_1(X) | Z = z_m)] = 0.2 \times 0.4 + 0.1 \times 0.2 + 0.7 \times 0.3 = 0.31;$$

$$\sum_{m=1}^3 [P(Z = z_m) \cdot P(P(X) = p_2(X) | Z = z_m)] = 0.2 \times 0.6 + 0.1 \times 0.8 + 0.7 \times 0.7 = 0.69;$$

Now we apply (15):

$$P(Y = Rich) = 0.7 \times 0.2 \times 0.31 + 0.4 \times 0.8 \times 0.31 + 0.7 \times 0.5 \times 0.69 + 0.4 \times 0.5 \times 0.69 = 0.5221;$$

$$P(Y = Poor) = 0.4779.$$

3 Bayesian Metanetwork for Managing Probability Distributions

We define a Bayesian Metanetwork in a way implementing the basic intuition we had defining a Semantic Metanetwork few years ago. Semantic Metanetwork [16] was defined as a set of semantic networks, which are put on each other in such a way that links of every previous semantic network are in the same time nodes of the next network. In a Semantic Metanetwork every higher level controls semantic structure of the lower level. Simple controlling rules might be, for example, in what contexts certain link of a semantic structure can exist and in what context it should be deleted from the semantic structure.

Definition. Bayesian Metanetwork is a set of Bayesian networks, which are put on each other in such a way that conditional or unconditional probability distributions associated with nodes of every previous probabilistic network depend on probability distributions associated with nodes of the next network.

First consider 2-level Bayesian Metanetwork (the idea is shown in Fig. 5). Context variables in it are considered to be on a second (contextual) level to control the conditional probabilities associated with predictive level of the network. Standard Bayesian inference is applied in Bayesian network of each level.

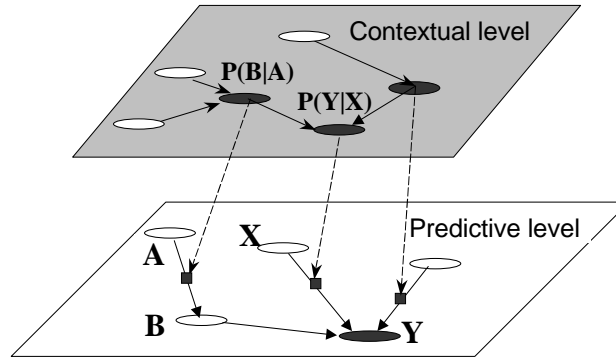


Fig. 5. Two-level Bayesian Metanetwork for managing conditional probabilities

A sample of a Bayesian Metanetwork (for simplicity projected to 2-D space), which is part of the metanetwork in Fig. 5, is presented in Fig. 6. Bayesian (meta)network in Fig. 6 has the following parameters:

Attributes of the predictive level:

A with values $\{a_1, a_2, \dots, a_{na}\}$; B with values $\{b_1, b_2, \dots, b_{nb}\}$ and probability $P(A)$;

X with values $\{x_1, x_2, \dots, x_{nx}\}$; Y with values $\{y_1, y_2, \dots, y_{ny}\}$ and probability $P(X)$;

Conditional probabilities of the predictive level: $P(B|A)$ which is random variable with set of values $\{p_1(B|A), p_2(B|A), \dots, p_{mp}(B|A)\}$. Important to notice that this parameter serves as an ordinary conditional probability in the predictive level of Bayesian metanetwork and it also serves as an attribute node on the contextual level.

$P(Y|X)$ which is random variable with possible values $\{p_1(Y|X), p_2(Y|X), \dots, p_{np}(Y|X)\}$ and also considered as an attribute node on a contextual level of Bayesian Metanetwork.

Conditional probability from the contextual level: $P(P(Y|X), P(B|A))$, which defines conditional probability between two contextual attributes $P(B|A)$ and $P(Y|X)$.

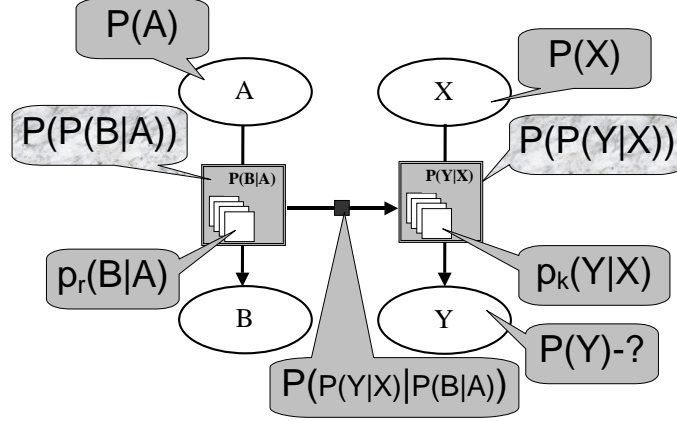


Fig. 6. An example of Bayesian Metanetwork. The nodes of the 2nd-level network correspond to the conditional probabilities of the 1st-level network $P(B|A)$ and $P(Y|X)$. The arc in the 2nd-level network corresponds to the conditional probability $P(P(Y|X)|P(B|A))$

The probability of the target attribute $P(Y)$ can be computed by applying basic Bayesian inference on both levels of the metanetwork. First we are exploring the basic level of the BMN. Finding joint probability:

$$\begin{aligned}
 & P(Y = y_j, X = x_i, P(Y | X) = p_k(Y | X)) = \\
 & = \underbrace{P(Y = y_j | X = x_i, P(Y | X) = p_k(Y | X))}_{p_k(Y=y_j|X=x_i)} \cdot \underbrace{P(X = x_i, P(Y | X) = p_k(Y | X))}_{independent} = \\
 & = p_k(Y = y_j | X = x_i) \cdot P(X = x_i) \cdot P(P(Y | X) = p_k(Y | X));
 \end{aligned}$$

Then making marginalization:

$$P(Y = y_j) = \sum_k \sum_i [p_k(Y = y_j | X = x_i) \cdot P(X = x_i) \cdot P(P(Y | X) = p_k(Y | X))]; \quad (17)$$

Now exploring the metalevel (joint probability first):

$$\begin{aligned}
 & P(P(Y | X) = p_k(Y | X), P(B | A) = p_r(B | A)) = \\
 & = P(P(Y | X) = p_k(Y | X) | P(B | A) = p_r(B | A)) \cdot P(P(B | A) = p_r(B | A));
 \end{aligned}$$

Then marginalization on the metalevel:

$$\begin{aligned}
 & P(P(Y | X) = p_k(Y | X)) = \\
 & = \sum_r [P(P(Y | X) = p_k(Y | X) | P(B | A) = p_r(B | A)) \cdot P(P(B | A) = p_r(B | A))]; \quad (18)
 \end{aligned}$$

Finally from (17) and (18) we are getting the target probability:

$$P(Y = y_j) = \sum_i \sum_k \{ p_k(Y = y_j | X = x_i) \cdot P(X = x_i) \times \\ \times \sum_r [P(P(Y | X) = p_k(Y | X) | P(P(B | A) = p_r(B | A)) \times P(P(B | A) = p_r(B | A))] \}$$

Similar inference as for above case can be also applied to other cases of a Metanetwork where unconditional, conditional or both probability distributions associated with nodes of predictive level of the metanetwork depend on probability distributions associated with nodes of the contextual level of the metanetwork (Fig. 7).

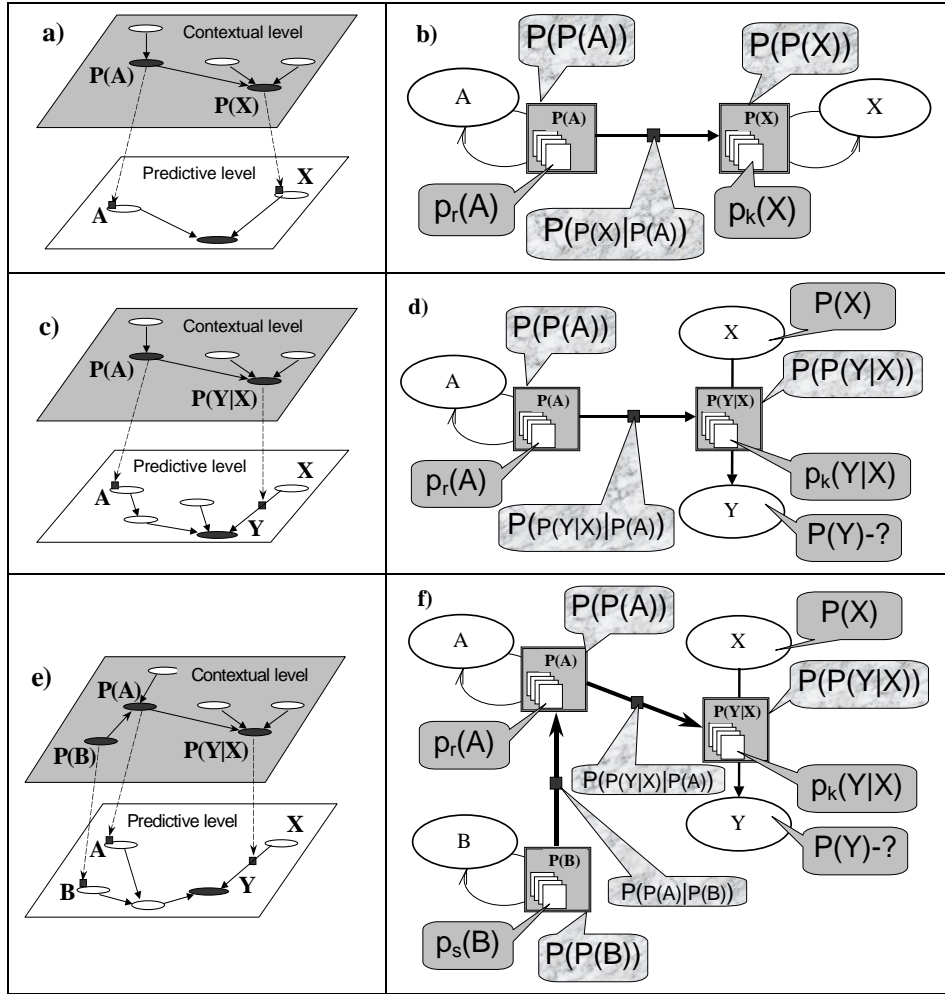


Fig. 7. Some other (than in Fig. 5) cases of Bayesian Metanetwork: In metanetwork (a) unconditional probability distributions associated with nodes of the predictive level network depend on probability distributions associated with nodes of the contextual level network, 2-D fragment is shown in (b); in (c) the contextual level metanetwork models conditional dependence particularly between unconditional and conditional probabilities of the predictive level (see also (d)); in (e) and (f) the combination of two previous cases is shown.

4 Conclusions

The main challenge of this paper is the multilevel probabilistic meta-model (Bayesian Metanetwork), which is an extension of traditional BN and modification of recursive multinet. The model assumes that interoperability between component networks can be also modeled by another BN. Bayesian Metanetwork is a set of BN, which are put on each other in such a way that conditional or unconditional probability distributions associated with nodes of every previous probabilistic network depend on probability distributions associated with nodes of the next network. We assume parameters (probability distributions) of a BN as random variables and allow conditional dependencies between these probabilities. Several cases of two-level Bayesian Metanetworks were presented, which consist on interrelated predictive and contextual BN models. By recursive application of the same “meta”-principle we can assume that a Bayesian metanetwork might have as many levels as necessary depending on the dynamics and complexity of the context. In this paper we have considered only some basic concepts, definitions and inference in a Bayesian Metanetwork. Some preliminary usage cases were discussed in [17], where context-sensitive mobile user preferences were modeled by Bayesian metanetworks. The basic algorithm for learning Bayesian metanetworks from data can be found in [18]. However further research is needed for providing advanced learning algorithms for such networks as well as for proving its efficiency on real-world examples.

Acknowledgements

I am grateful for Dr. Oleksandra Vitko for fruitful discussions and valuable comments within the scope of this paper. Also I appreciate anonymous reviewers for their useful feedback and concrete suggestions to improve the paper. The partial support for this research, provided by TEKES (SmartResource project) and cooperating companies: TeliaSonera, Metso Automation, TietoEnator, Science Park, Agora Center and University of Jyväskylä, was also highly appreciated.

References

1. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, (Morgan Kaufmann, 1988).
2. M. Henrion, Some Practical Issues in Constructing Belief Networks, In: *Proceedings of the 3-rd Annual Conference on Uncertainty in Artificial Intelligence*, (Elsevier, 1989), pp. 161-174.
3. D. Heckerman, A Tutorial on Learning with Bayesian Networks, *Technical Report MSR-TR-95-06*, (Microsoft Research, March 1995).
4. C. J. Butz, Exploiting Contextual Independencies in Web Search and User Profiling, In: *Proceedings of the World Congress on Computational Intelligence*, (Hawaii, USA, 2002), pp. 1051-1056.

5. C. Boutilier, N. Friedman, M. Goldszmidt and D. Koller, Context-Specific Independence in Bayesian Networks, In: *Proceedings of the 12-th Conference on Uncertainty in Artificial Intelligence*, (Portland, USA, 1996), pp. 115-123.
6. N.L. Zhang, Inference in Bayesian networks: The Role of Context-Specific Independence, *International Journal of Information Technology and Decision Making*, **1**(1) 2002, 91-119.
7. D. Geiger and D. Heckerman, Knowledge Representation and Inference in Similarity Networks and Bayesian Multinets, *Artificial Intelligence*, Vol. 82, (Elsevier, 1996), pp. 45-74.
8. E. M. Helsper and L. C. van der Gaag, Building Bayesian Networks through Ontologies, In: *Proceedings of the 15th European Conference on Artificial Intelligence*, ed. F. van Harmelen, (IOS Press, Amsterdam, 2002), pp. 680-684.
9. J.-W. Bang, R. Chaleil and J. Williamson, Two-Stage Bayesian Networks for Metabolic Network Prediction, In: *Proceedings of the Workshop on Qualitative and Model-Based Reasoning in Biomedicine*, ed. P. Lucas (Protaras, Cyprus, 2003), pp. 19-23.
10. A. Singhal, J. Luo and C. Brown, A Multi-Level Bayesian Network Approach to Image Sensor Fusion, In: *Proceedings of FUSION 2000: 3rd International Conference on Information Fusion*, Vol. 2, (IEEE, 2000), pp. 9-16.
11. K. W. Przytula and D. Thompson, Construction of Bayesian Networks for Diagnostics, In: *Proceedings of the 2000 IEEE Aerospace Conference*, (IEEE Aerospace and Electronics Systems Society, 2000).
12. N. Friedman, D. Geiger, and M. Goldszmidt, Bayesian Network Classifiers, *Machine Learning*, **29**(2-3), (Kluwer, 1997), pp. 131-161.
13. J. Cheng and R. Greiner, Learning Bayesian Belief Network Classifiers: Algorithms and System, In: *Proceedings of the 14-th Canadian Conference on Artificial Intelligence*, Lecture Notes in Computer Science, Vol. 2056, (Springer-Verlag Heidelberg, 2001), pp. 141-151.
14. J. A. Bilmes, Dynamic Bayesian Multinets, In: *Proceedings of the 16-th Conference on Uncertainty in Artificial Intelligence*, (Morgan Kaufmann, San Francisco, 2000), pp. 38-45.
15. J. Pena, J. A. Lozano, and P. Larranaga, Learning Bayesian Networks for Clustering by Means of Constructive Induction, *Machine Learning*, **47**(1), (Kluwer, 2002), pp. 63-90.
16. V. Terziyan and S. Puuronen, Reasoning with Multilevel Contexts in Semantic Metanetworks, In: *Formal Aspects in Context*, eds. P. Bonzon, M. Cavalcanti, R. Nossun (Kluwer, 2000), pp. 107-126.
17. V. Terziyan and O. Vitko, Bayesian Metanetwork for Modelling User Preferences in Mobile Environment, In: *Proceedings of KI 2003: Advances in Artificial Intelligence*, Lecture Notes in Artificial Intelligence, Vol. 2821, ed. A. Gunter, R. Kruse and B. Neumann, (Springer-Verlag, 2003), pp.370-384.
18. V. Terziyan and O. Vitko, Learning Bayesian Metanetworks from Data with Multilevel Uncertainty, In: *Proceedings of the First IFIP International Conference on Artificial Intelligence and Innovations (AIAI-2004)*, Toulouse, France, (Kluwer, 2004), to appear.