

Bayesian Metanetwork for Context-Sensitive Feature Relevance

Vagan Terziyan

Industrial Ontologies Group, Agora Center, University of Jyväskylä,
P.O. Box 35 (Agora), FIN-40014 Jyväskylä, Finland
vagan@it.jyu.fi

Abstract. Bayesian Networks are proven to be a comprehensive model to describe causal relationships among domain attributes with probabilistic measure of appropriate conditional dependency. However, depending on task and context, many attributes of the model might not be relevant. If a network has been learned across multiple contexts then all uncovered conditional dependencies are averaged over all contexts and cannot guarantee high predictive accuracy when applied to a concrete case. We are considering a context as a set of contextual attributes, which are not directly effect probability distribution of the target attributes, but they effect on a “relevance” of the predictive attributes towards target attributes. In this paper we use the Bayesian Metanetwork vision to model such context-sensitive feature relevance. Such model assumes that the relevance of predictive attributes in a Bayesian network might be a random attribute itself and it provides a tool to reason based not only on probabilities of predictive attributes but also on their relevancies. According to this model, the evidence observed about contextual attributes is used to extract a relevant substructure from a Bayesian network model and then the predictive attributes evidence is used to reason about probability distribution of the target attribute in the extracted sub-network. We provide the basic architecture for such Bayesian Metanetwork, basic reasoning formalism and some examples.

1 Introduction

A *Bayesian network* is a valuable tool for reasoning about probabilistic (causal) relationships [1]. A Bayesian network for a set of attributes $X = \{X_1, \dots, X_n\}$ is a directed acyclic graph with a network structure S that encodes a set of conditional independence assertions about attributes in X , and a set P of local probability distributions associated with each attribute [2].

An important task in learning Bayesian networks from data is model selection [3]. The models-candidates are evaluated according to measured degree to which a network structure fits the prior knowledge and data. Then the best structure is selected or several good structures are processed in model averaging. Each attribute in ordinary Bayesian network has the same status, so they are just combined in possible models-candidates to encode possible conditional dependencies however many modifications of Bayesian networks 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 and which influences the probability distribution of the target attribute(s).
- *Contextual attribute*, which has not direct visible effect to target attributes but influences relevance of attributes in the predictive model. A contextual attribute can be conditionally dependent on some other contextual attribute.

Causal independence in a Bayesian network refers to the situation where multiple causes provided by predictive attributes contribute independently to a common effect on a target attribute. Context specific independence refers to such dependencies that depend on particular values of contextual attributes. 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 Bayesian networks. Boutilier et al. [5] presents two algorithms to exploit context specific independence in a Bayesian network. The first one is network transformation and clustering. The other one is a form of cutset conditioning using reasoning by cases, where each case is a possible assignment to the variables in the cutset. Zhang [6] presents a rule-based contextual variable elimination algorithm. Contextual variable elimination represents conditional probabilities in terms of generalized rules, which capture context specific independence in variables. Geiger and Heckerman [7] present method to exploit context specific independence. With the notion of similarity networks, context specific independencies are made explicit in the graphical structure of a Bayesian network. Bayesian Multi-nets were first introduced in ([8]) and then studied in ([9]) as a type of classifiers 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. A recursive Bayesian multinet was introduced by Pena et al [10] as a decision tree with component Bayesian networks at the leaves. The idea was to decompose the learning problem into learning component networks from incomplete data.

In our previous work [11, 12], is the multilevel probabilistic meta-model (Bayesian Metanetwork), has been presented, which is an extension of traditional BN and modification of recursive multinets. It assumes that interoperability among 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. Algorithms for learning Bayesian Metanetworks were discussed in [13].

As our main goal in this paper, we are presenting another view to the Bayesian Metanetwork by presenting the concept of attribute “relevance” as additional (to an attribute value probability) computational parameter of a Bayesian Network. Based on computed relevance only a specific sub-network from the whole Bayesian Network will be extracted and used for reasoning. The rest of paper organized as follows. In Section 2 we first provide basic architecture of the Bayesian Metanetwork for

managing Attribute Relevance. In Section 3 we provide the reasoning formalism and few examples. A general concept of a Relevance Metanetwork is given in Section 4. We conclude in Section 5.

2 Bayesian Metanetwork for Managing Attributes' Relevance

Relevance is a property of an attribute as a whole, not a property of certain values of an attribute (see Fig. 1). This makes a difference between relevance and probability, because the last one has as many values as an attribute itself. Another words, when we say probability, we mean probability of the value of the attribute, when we say relevance, we mean relevance (*probability to be included to the model*) of the attribute as whole.

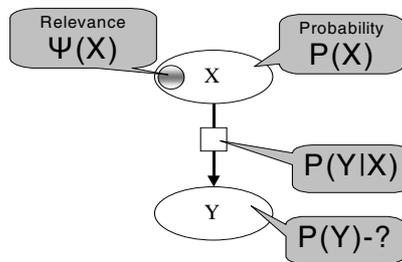


Fig. 1. The relevance of an attribute in a Bayesian Network

Bayesian Network in Fig. 1 actually includes two following subnetworks (see Fig. 2), which illustrate the definition of a “relevance”.

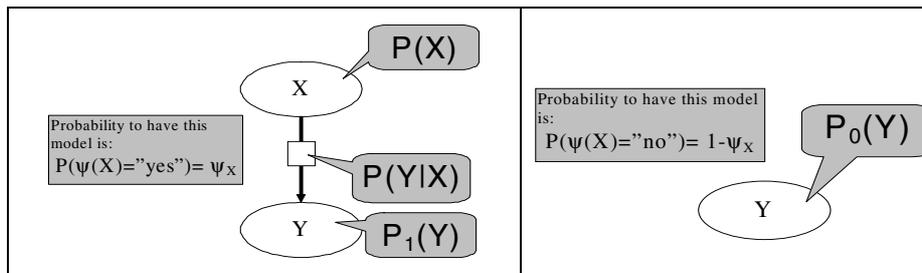


Fig. 2. Two valid subnetworks based on the relevance of the attribute X

In the network from Fig. 1 and Fig. 2 we have: (a) attributes: predictive attribute X with values $\{x_1, x_2, \dots, x_{nx}\}$, target attribute Y with values $\{y_1, y_2, \dots, y_{ny}\}$; (b) probability distributions of the attributes: $P(X)$, $P(Y|X)$; (c) posteriori probability distributions of the target attribute: $P_1(Y)$ and $P_0(Y)$ for the two cases (X – relevant or X irrelevant) of the valid subnetwork respectively. Relevance predicate:

$\psi(X)$ = “yes”, if parameter X is relevant; $\psi(X)$ = “no”, if parameter X is not relevant.

Relevance value: $\psi_X = P(\psi(X) = \text{“yes”})$.

Let’s estimate $P(Y)$ based on Bayesian reasoning:

$$P(Y) = \psi_X \cdot P_1(Y) + (1 - \psi_X) \cdot P_0(Y), \quad \text{where} \quad (1)$$

$$P_1(Y) = \sum_{i=1}^{nx} P(Y | X = x_i) \cdot P(X = x_i). \quad (2)$$

$P_0(Y)$ can be calculated based on $P(Y|X)$ knowing that in that case Y is not depending on X , because X is considered as not relevant:

$$P_0(Y) = \frac{1}{nx} \sum_{i=1}^{nx} P(Y | X = x_i). \quad (3)$$

Substituting (2) and (3) to (1) we obtain:

$$\begin{aligned} P(Y) &= \psi_X \cdot \sum_{i=1}^{nx} [P(Y | X = x_i) \cdot P(X = x_i)] + (1 - \psi_X) \cdot \frac{1}{nx} \cdot \sum_{i=1}^{nx} P(Y | X = x_i) = \\ &= \sum_{i=1}^{nx} P(Y | X = x_i) \cdot [\psi_X \cdot P(X = x_i) + \frac{(1 - \psi_X)}{nx}], \end{aligned}$$

$$\text{which is in compact form is: } P(Y) = \frac{1}{nx} \cdot \sum_X P(Y | X) \cdot [nx \cdot \psi_X \cdot P(X) + (1 - \psi_X)]. \quad (4)$$

Consider example, where the attribute X will be “state of whether” and attribute Y , which is influenced by X , will be “state of mood”. Let the values of the attributes and appropriate prior probabilities will be as follows:

X (“state of weather”) = {“sunny”, “overcast”, “rain”}; Y (“state of mood”) = {“good”, “bad”};
 $P(X=\text{“sunny”}) = 0.4$; $P(X=\text{“overcast”}) = 0.5$; $P(X=\text{“rain”}) = 0.1$;
 $P(Y=\text{“good”}|X=\text{“sunny”}) = 0.7$; $P(Y=\text{“good”}|X=\text{“overcast”}) = 0.5$; $P(Y=\text{“good”}|X=\text{“rain”}) = 0.2$;

Let conditional probability, which links X and Y will be as follows:

$P(Y=\text{“bad”}|X=\text{“sunny”}) = 0.3$; $P(Y=\text{“bad”}|X=\text{“overcast”}) = 0.5$; $P(Y=\text{“bad”}|X=\text{“rain”}) = 0.8$;

Assume the value of relevance for the attribute X is known and equal: $\psi_X = 0.6$.

Now, according to (4) we have:

$$\begin{aligned} P(Y = \text{“good”}) &= \frac{1}{3} \cdot \{P(Y = \text{“good”} | X = \text{“sunny”}) \cdot [1.8 \cdot P(X = \text{“sunny”}) + 0.4] + \\ &+ P(Y = \text{“good”} | X = \text{“overcast”}) \cdot [1.8 \cdot P(X = \text{“overcast”}) + 0.4] + \\ &+ P(Y = \text{“good”} | X = \text{“rain”}) \cdot [1.8 \cdot P(X = \text{“rain”}) + 0.4]\} = 0.517; \end{aligned}$$

Similarly: $P(Y = \text{“bad”}) = 0.483$.

One can also notice that these values belong to the intervals created by the two extreme cases, when parameter X is not relevant at all or it is fully relevant:

$$0.467 \approx P_0(Y = \text{"good"})|_{\psi_x=0} < P(Y = \text{"good"})|_{\psi_x=0.6} < P_1(Y = \text{"good"})|_{\psi_x=1} = 0.55 ;$$

$$0.45 = P_1(Y = \text{"bad"})|_{\psi_x=1} < P(Y = \text{"bad"})|_{\psi_x=0.6} < P_0(Y = \text{"bad"})|_{\psi_x=0} \approx 0.533 .$$

3 General Formalism and Samples

More complicated case is the management of relevance in the following situation (Fig. 3):

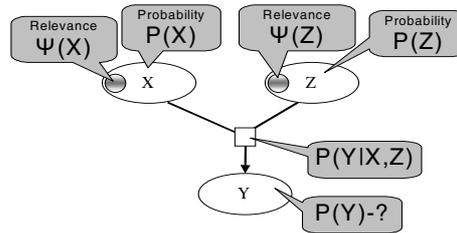


Fig. 3. Relevance management with two predictive attributes

Here we have 4 following subnetworks depending on the relevance (see Fig. 4).

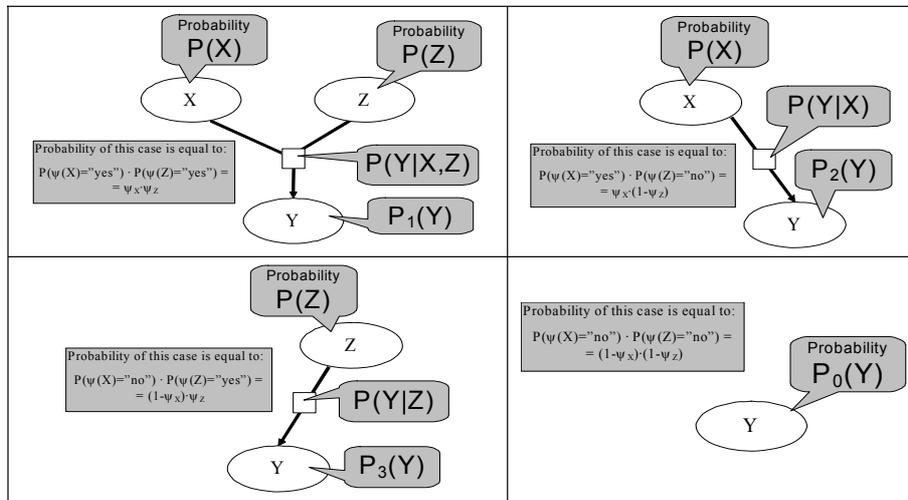


Fig. 4. Subnetworks for the case with two predictive attributes

Here we have: (a) predictive attributes: X with values $\{x_1, x_2, \dots, x_{nx}\}$; Z with values $\{z_1, z_2, \dots, z_{nz}\}$; (b) target attribute Y with values $\{y_1, y_2, \dots, y_{ny}\}$; (c) probabilities: P(X), P(Z), P(Y|X,Z); (d) relevance predicate: $\psi(X) = \text{"yes"}$, if X is relevant; $\psi(X) = \text{"no"}$, if X is not relevant; (e) relevancies: $\psi_x = P(\psi(X) = \text{"yes"})$; $\psi_z = P(\psi(Z) = \text{"yes"})$.

Let's estimate $P(Y)$:

$$P(Y) = \psi_X \cdot \psi_Z \cdot P_1(Y) + \psi_X \cdot (1 - \psi_Z) \cdot P_2(Y) + (1 - \psi_X) \cdot \psi_Z \cdot P_3(Y) + (1 - \psi_X) \cdot (1 - \psi_Z) \cdot P_0(Y) \quad (5)$$

$$P_1(Y) = \sum_{i=1}^{nx} \sum_{k=1}^{nz} P(Y | X = x_i, Z = z_k) \cdot P(X = x_i) \cdot P(Z = z_k). \quad (6)$$

$$P_2(Y) = \sum_{i=1}^{nx} P(Y | X = x_i) \cdot P(X = x_i). \quad (7)$$

$$P_3(Y) = \sum_{k=1}^{nz} P(Y | Z = z_k) \cdot P(Z = z_k). \quad (8)$$

Now we should extract $P(Y|X)$, $P(Y|Z)$, $P_0(Y)$ from $P(Y|X,Z)$ and $P(Y|Z)$ from $P(Y|X,Z)$, which is:

$$P(Y | X) = \frac{1}{nz} \cdot \sum_{k=1}^{nz} P(Y | X, Z = z_k), \quad (9)$$

$$P(Y | Z) = \frac{1}{nx} \cdot \sum_{i=1}^{nx} P(Y | X = x_i, Z), \quad (10)$$

$$P_0(Y) = \frac{1}{nx \cdot nz} \cdot \sum_{i=1}^{nx} \sum_{k=1}^{nz} P(Y | X = x_i, Z = z_k), \quad (11)$$

We can rewrite (7) using (9) as follows:

$$P_2(Y) = \frac{1}{nz} \cdot \sum_{i=1}^{nx} \sum_{k=1}^{nz} P(Y | X = x_i, Z = z_k) \cdot P(X = x_i). \quad (12)$$

$$\text{From (8) and (10): } P_3(Y) = \frac{1}{nx} \cdot \sum_{i=1}^{nx} \sum_{k=1}^{nz} P(Y | X = x_i, Z = z_k) \cdot P(Z = z_k). \quad (13)$$

Finally we can substitute (6), (11), (12), (13) to (5):

$$\begin{aligned} P(Y) &= \psi_X \cdot \psi_Z \cdot \sum_{i=1}^{nx} \sum_{k=1}^{nz} P(Y | X = x_i, Z = z_k) \cdot P(X = x_i) \cdot P(Z = z_k) + \psi_X \cdot (1 - \psi_Z) \cdot \frac{1}{nz} \cdot \\ &\cdot \sum_{i=1}^{nx} \sum_{k=1}^{nz} P(Y | X = x_i, Z = z_k) \cdot P(X = x_i) + (1 - \psi_X) \cdot \psi_Z \cdot \frac{1}{nx} \cdot \sum_{i=1}^{nx} \sum_{k=1}^{nz} P(Y | X = x_i, Z = z_k) \cdot P(Z = z_k) + \\ &+ (1 - \psi_X) \cdot (1 - \psi_Z) \cdot \frac{1}{nx \cdot nz} \cdot \sum_{i=1}^{nx} \sum_{k=1}^{nz} P(Y | X = x_i, Z = z_k), \end{aligned}$$

which is in a more compact form:

$$\begin{aligned} P(Y) &= \frac{1}{nx \cdot nz} \cdot \sum_X \sum_Z P(Y | X, Z) \cdot [nx \cdot nz \cdot \psi_X \cdot \psi_Z \cdot P(X) \cdot P(Z) + \\ &+ nx \cdot \psi_X \cdot (1 - \psi_Z) \cdot P(X) + nz \cdot (1 - \psi_X) \cdot \psi_Z \cdot P(Z) + (1 - \psi_X) \cdot (1 - \psi_Z)]. \end{aligned} \quad (14)$$

Consider example. Let we have the following set of data:

X	Z	Y
Sunny	Alone	Good
Overcast	With girlfriend	Bad
Overcast	With dog	Bad
Sunny	With dog	Good
Sunny	Alone	Good
Overcast	Alone	Bad
Rain	With girlfriend	Bad
Sunny	With dog	Good
Overcast	With dog	Bad
Sunny	With girlfriend	Bad
Overcast	With girlfriend	Good
Overcast	Alone	Bad
Overcast	With dog	Bad
Sunny	With girlfriend	Good
Overcast	With dog	Bad
Overcast	Alone	Bad
Sunny	Alone	Bad
Sunny	With dog	Bad
Rain	With girlfriend	Good
Overcast	With dog	Good

X (“state of weather”) = {“sunny”, “overcast”, “rain”}
 Z (“companion”) = {“alone”, “girlfriend”, “dog”}
 Y (“state of mood”) = {“good”, “bad”};

$P(X=“sunny”) = 0.4;$ $P(X=“overcast”) = 0.5;$ $P(X=“rain”) = 0.1;$
 $P(Z=“alone”) = 0.3;$ $P(Z=“girlfriend”) = 0.3;$ $P(Z=“dog”) = 0.4;$

P(Y=“good” X, Z)	Z = “alone”	Z=“girlfriend”	Z=“dog”
X = “sunny”	0.667	0.5	0.667
X = “overcast”	0	0.5	0.25
X = “rain”	0	0.5	0
P(Y=“bad” X, Z)	Z = “alone”	Z=“girlfriend”	Z=“dog”
X = “sunny”	0.333	0.5	0.333
X = “overcast”	1	0.5	0.75
X = “rain”	1	0.5	1

According to (9):

P(Y X)	X = “sunny”	X = “overcast”	X = “rain”
Y = “good”	0.611	0.25	0.167
Y = “bad”	0.389	0.75	0.833

According to (10):

P(Y Z)	Z = “alone”	Z = “girlfriend”	Z = “dog”
Y = “good”	0.222	0.5	0.306
Y = “bad”	0.778	0.5	0.694

According to (11): $P_0(Y="good") = 0.3426$; $P_0(Y="bad") = 0.6574$.

Assuming that relevancies of our parameters are as follows: $\psi_X = 0.8$, $\psi_Z = 0.5$, we can estimate $P(Y)$ based on (14): $P(Y="good") \approx 0.3773$; $P(Y="bad") \approx 0.6227$.

One can also notice that these values belong to the interval created by the two extreme cases, when parameters are not relevant at all or they are fully relevant:

$$0.3426 \approx P_0(Y="good")|_{\psi_X=0, \psi_Z=0} < P(Y="good")|_{\psi_X=0.8, \psi_Z=0.5} < P_1(Y="good")|_{\psi_X=1, \psi_Z=1} \approx 0.3867 ;$$

$$0.6133 \approx P_1(Y="bad")|_{\psi_X=1, \psi_Z=1} < P(Y="bad")|_{\psi_X=0.8, \psi_Z=0.5} < P_0(Y="bad")|_{\psi_X=0, \psi_Z=0} \approx 0.6574.$$

Consider the general case of managing relevance (Fig. 5):

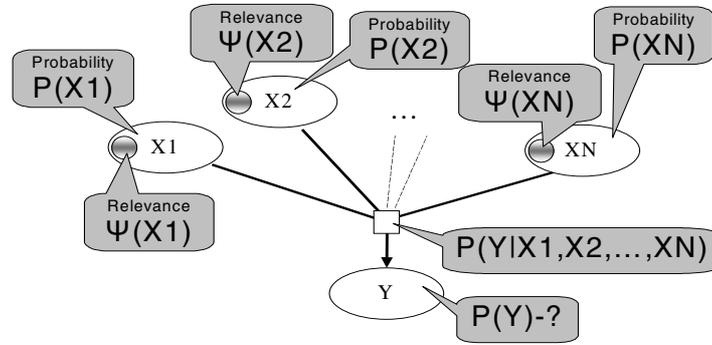


Fig. 5. General case of relevance management

In this case we have the following: (a) predictive attributes: X_1 with values $\{x_{11}, x_{12}, \dots, x_{1n_{x1}}\}$; X_2 with values $\{x_{21}, x_{22}, \dots, x_{2n_{x2}}\}$; ... X_N with values $\{x_{n1}, x_{n2}, \dots, x_{n_{n_{xn}}}\}$; (b) target attribute: Y with values $\{y_1, y_2, \dots, y_{ny}\}$; (c) probabilities: $P(X_1)$, $P(X_2), \dots, P(X_N)$; $P(Y|X_1, X_2, \dots, X_N)$; (d) relevancies: $\psi_{X_1} = P(\psi(X_1) = \text{"yes"})$; $\psi_{X_2} = P(\psi(X_2) = \text{"yes"})$; ... $\psi_{X_N} = P(\psi(X_N) = \text{"yes"})$. Task is to estimate $P(Y)$.

Generalizing (4) and (14) to the case of N predictive variables we finally obtain:

$$P(Y) = \frac{1}{\prod_{s=1}^N n_{xs}} \cdot \sum_{X_1} \sum_{X_2} \dots \sum_{X_N} [P(Y | X_1, X_2, \dots, X_N) \cdot \prod_{\forall r(\psi(X_r)=\text{"yes"})} n_{xr} \cdot \psi_{X_r} \cdot P(X_r) \cdot \prod_{\forall q(\psi(X_q)=\text{"no"})} (1 - \psi_{X_q})]$$

4 A Relevance Metanetwork

Relevance Bayesian Metanetwork can be defined on a given predictive probabilistic network (Fig. 6). It encodes the conditional dependencies over the relevancies. Relevance metanetwork contains prior relevancies and conditional relevancies.

Considering such definition of relevance metanetwork over the predictive network it is clear that the strict correspondence between nodes of both network exists but the arcs do not need to be strictly corresponding (as shown in Fig. 6). It means that relevancies of two variables can be dependent, although their values are conditionally independent and vice versa (Fig. 7). So, the topologies of the networks are different in general case.

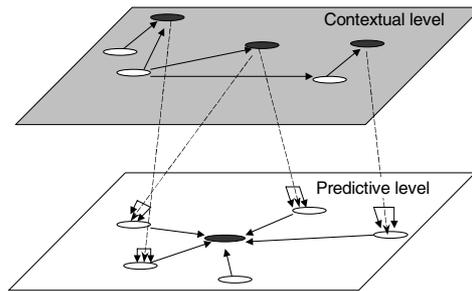


Fig. 6. Relevance network defined over the predictive network

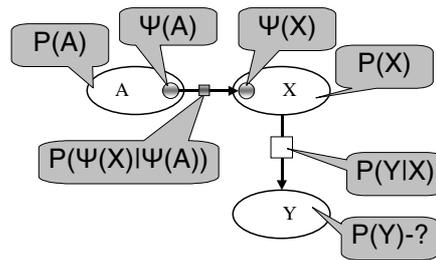


Fig. 7. Architecture of a simple relevance metanetwork

In a relevance network the relevancies are considered as random variables between which the conditional dependencies can be learned. For example in Fig. 7, the probability of target attribute Y can be computed as follows:

$$P(Y) = \frac{1}{nx} \cdot \sum_X \{P(Y | X) \cdot [nx \cdot P(X) \cdot \sum_{\psi_A} P(\psi_X | \psi_A) \cdot P(\psi_A) + (1 - \psi_X)]\}.$$

More complicated example of a Bayesian metanetwork completed from predictive and relevance networks is shown in Fig. 8. The graph of the Metanetwork in the figure consists of two subgraphs (a) predictive network layer and (b) relevance network layer. The challenge here is that the relevance network subgraph models the relevance conditional dependency and in the same time the posteriori relevance values calculated with this graph effect the calculations at the basic predictive subgraph.

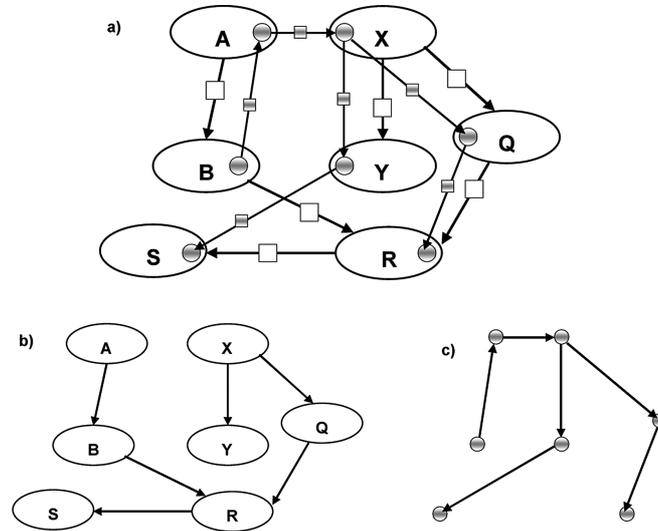


Fig. 8. Example of Bayesian metanetwork (a), consisting of the predictive subnetwork (b) and relevance subnetwork (c) both of which have corresponding nodes, but different topologies

5 Conclusions

Bayesian Networks are proven to be a comprehensive model to describe causal relationships among domain attributes with probabilistic measure of appropriate conditional dependency. However, depending on task and context, many attributes of the model might not be relevant. If a Bayesian Network has been learned across multiple contexts then all uncovered conditional dependencies are averaged over all contexts and cannot guarantee high predictive accuracy when applied to a concrete case. We are considering a context as a set of contextual attributes, which are not directly effect probability distribution of the target attributes, but they effect on a “relevance” of the predictive attributes towards target attributes. In this paper we use the Bayesian Metanetwork vision to model such context-sensitive feature relevance. Such model assumes that the relevance of predictive attributes in a Bayesian network might be a random attribute itself and it provides a tool to reason based not only on probabilities of predictive attributes but also on their relevancies. According to this model, the evidence observed about contextual attributes is used to extract a relevant substructure from a Bayesian network model and then the predictive attributes evidence is used to reason about probability distribution of the target attribute in the extracted sub-network. Such models are useful when the relevance of the attributes essentially depends on the context.

Acknowledgment

This research has been performed as a part of the SmartResource (“Proactive Self-Maintained Resources in Semantic Web”) project of Agora Center (University of Jyväskylä, Finland) and funded by TEKES and industrial consortium of following companies: Metso Automation, TeliaSonera, TietoEnator and Science Park.

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: *Proc. 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. N. Friedman, D. Geiger, and M. Goldszmidt, Bayesian Network Classifiers, *Machine Learning*, **29**(2-3), (Kluwer, 1997), pp. 131-161.
9. 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.
10. 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.
11. V. Terziyan, A Bayesian Metanetwork, *International Journal on Artificial Intelligence Tools*, **14**(3), (World Scientific, 2005), pp. 371-384.
12. 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.
13. Terziyan and O. Vitko, Learning Bayesian Metanetworks from Data with Multilevel Uncertainty, In: M. Bramer and V. Devedzic (eds.), *Proc. of the First IFIP International Conf. on Artificial Intelligence and Innovations*, Toulouse, France, (Kluwer, 2004), pp. 187-196.