LEARNING BAYESIAN METANETWORKS FROM DATA WITH MULTILEVEL UNCERTAINTY

Vagan Terziyan, Oleksandra Vitko
University of Jyväskylä,P.O. Box 35 (Agora), FIN-40014 Jyväskylä, Finland, vagan@it.jyu.fi
Kharkov National University of Radioelectronics, Lenin Avenue 14, 61166 Kharkov, Ukraine, vitko@kture.kharkov.ua

Abstract: Managing knowledge by maintaining it according to dynamic context is among the basic abilities of a knowledge-based system. The two main challenges in managing context in Bayesian networks are the introduction of contextual (in)dependence and Bayesian multinets. We are presenting one possible implementation of a context sensitive Bayesian multinet – the Bayesian Metanetwork, which implies that interoperability between component Bayesian networks (valid in different contexts) can be also modelled by another Bayesian network. The general concepts and two kinds of such Metanetwork models are considered. The main focus of this paper is learning procedure for Bayesian Metanetworks.

Key words: Bayesian Metanetworks, Bayesian learning, multilevel uncertainty, context

1. INTRODUCTION

Creating and managing knowledge according to different levels of possible context – are among the basic abilities of an intelligent system. Multilevel representation of a context allows reasoning with contexts towards solution of the following problems [9]:

- to derive knowledge interpreted using all known levels of its context;
- to derive unknown knowledge when interpretation of it in some context and the context itself are known;

- to derive unknown knowledge about a context when it is known how the knowledge is interpreted in this context;
- to transform knowledge from one context to another one.

Metanetwork-based models (e.g. the Semantic Metanetworks, the MetaPetrinets, etc.) have proved to be more powerful tools for knowledge representation in the presence of multiple contexts [8, 9].

A Bayesian network is known to be a valuable tool for encoding, learning and reasoning about probabilistic (casual) relationships. The Bayesian network for a set of variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ is a directed acyclic graph with the network structure **S** that encodes a set of conditional independence assertions about variables in **X**, and the set **P** of local probability distributions associated with each variable [4].

The two main challenges in utilizing context in Bayesian networks are the introduction of contextual independence [1] and Bayesian multinets. A recursive Bayesian multinet was introduced by Pena et al. [6] 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.

The main goal this research is to study another multiple Bayesian model – the Bayesian Metanetwork, which implies that interoperability between component Bayesian networks (valid in different contexts) can be also modelled by another Bayesian network. Such models suit well to e.g. user profiling applications where different probabilistic interrelations within predictive features from user's profile can be controlled by probabilistic interrelations among the contextual features, and other applications that require the formalism to manage two-level uncertainty or even multilevel uncertainty. The combination of the ideas of Metamodels and Bayesian network resulted to a refined and powerful formalism of a Bayesian Metanetwork.

The rest of the paper is organised as follows. In Section 2 we briefly introduce the formalism of Bayesian Metanetwork. In Section 3 we suggest the learning procedure for Bayesian Metanetwork. We conclude in Section 4.

2. THE BAYESIAN METANETWORKS

In our previous work [10], Bayesian Metanetwork formalism was used to model user preferences in mobile electronic commerce. Specific features and constraints of the mobile commerce environment demand the new flexible models of knowledge management. Such models assume to deal with the causal probabilistic relations in the cases when changes of a context occur. It was also shown that the Bayesian Metanetwork provides enough flexibility to be a powerful formalism also for many other data mining tasks [12].

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

The Bayesian Metanetwork is a triplet: $MBN = (BN, R, P)$, where BN = $\{BN_1, BN_2, \ldots BN_n\}$ is a set of Bayesian networks, each of which is considered on the appropriate level according to the index; R = $\{R_{1,2}, R_{2,3} \ldots R_{n-1,n}\}$ is a set of sets of interlevel links; P is a joint probability distribution over the Metanetwork.

Each $R_{i,i+1}$ is a set of interlevel links between *i* and *i+1* levels. We have proposed 2 types of links:

– $R_{v\text{-}e}$ is a link "vertex-edge" meaning that stochastic values of vertex v_{ik} in the network BN_i correspond to the different conditional probability tables $P_k(v_{i-1,j} \mid v_{i-1,pj})$ in the network BN_{i-1};
– $R_{v\text{-}v}$ is a link "vertex-vertex" meaning that stochastic values of vertex v_{ir} in the network BN_i correspond to the different relevance values of vertex $v_{i-1,r}$ in the network $BN_{i-1.}$

According to the introduced two types of interlevel links we consider two models of the Bayesian Metanetwork:

– *C*-Metanetwork, which has interlevel links of $R_{v\text{-}e}$ type used for managing conditional dependencies (**C**onditional Dependencies Metanetwork);
– *R*-Metanetwork, which has interlevel links of $R_{v\text{-}v}$ type used for modelling relevant feature selection (**R**elevance Metanetwork).

2.1 Bayesian C-Metanetwork for Managing Conditional Dependencies

In a C-Metanetwork the context variables are considered to be on the second (contextual) level to manage the conditional probabilities associated with the predictive level of the network [10, 12]. The sample of C-Metanetwork projected to 2-D space is presented in Figure 1.

Standard Bayesian inference is applied in the Bayesian network of each level. The examples and rules of propagation through the whole C-Metanetwork we have presented in [10, 12].

The two-level Metanetwork can be easily extended to the multilevel (multicontext) Metanetwork. In principle, we can assume that a Bayesian Metanetwork may have as many levels as necessary.

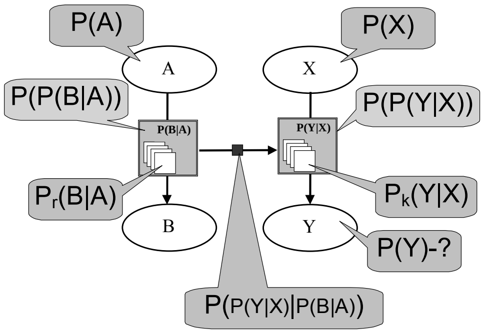

Figure 1. The sample of a Bayesian C-Metanetwork. The nodes of the 2nd-level network correspond to the conditional probabilities of the 1st-level network P(B|A) and P(Y|X). The directed arc in the 2nd-level network corresponds to the conditional probability P(P(Y|X)|P(B|A))

2.2 The Bayesian R-Metanetwork for Modelling Relevant Feature Selection

The Bayesian Metanetwork can be also used as a tool for the relevant feature selection. In R-Metanetwork the context variables are again considered as the higher-level control upon the basic network with predictive variables [10, 12]. Values of the context variables are assumed to have an influence to the relevancies of the variables on the predictive level.

We consider relevance value as a probability of importance of the variable to the inference of target attribute in the given context.

Contextual relevance network can be defined over the given predictive probabilistic network as it is shown in Figure 2.

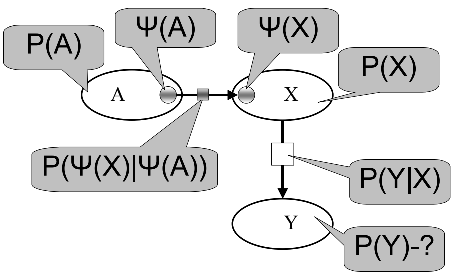

Figure 2. The simple relevance network with the attributes ψ(A), ψ(X) and the conditional probability P(ψ(X)| ψ(A)) defined over the predictive network with the attributes A, X, Y and the conditional probability P(Y|X)

The Bayesian R-Metanetwork in Figure 2 encodes the conditional dependencies over the relevancies and contains:

– the relevance predicate: $\psi(X)$ = "yes", if parameter X is relevant; $\psi(X)$ = "no", if parameter X is not relevant;
– the relevance value: $\psi_X = P(\psi(X) = \text{"yes"})$.

Standard Bayesian inference is applied to the Bayesian network of each level. The examples and rules of propagation through the whole R-Metanetwork we have presented in [10, 12].

3. LEARNING BAYESIAN METANETWORKS

In this Section we suggest the learning procedures for Bayesian Metanetworks. Both structure learning and parameter learning are considered.

A number of methods for learning a Bayesian network were developed and are in use, see e.g. [2, 4, 5]. Such methods can be applied for learning component Bayesian networks on each level of the Metanetwork. The main challenge of this work was the extension of the standard learning procedures for the case of multilevel probabilistic Metanetworks to enable learning interlevel relationships.

Let's consider the following learning task:

Given training set ***D*** of training examples $<X_1, X_2, \ldots X_n, Y>$,

Goal to restore:

– the set of levels of Bayesian Metanetwork $\{l_1, l_2, \ldots l_L\}$, each level is a Bayesian network;
– the interlevel links for each pair of successive levels $\{l_r, l_{r+1}\}$;
– the network structure and parameters at each level, particularly probabilities $P(v_i)$ and $P(v_i|parents(v_i))$ for each variable v_i.

We suggest the following learning procedure for Bayesian Metanetworks consisting of four stages; the last three of them are iteratively repeated at each level of the Metanetwork.

Stage 1. Division of attributes among the levels. The task of this stage is to divide the input vector of attributes $<X_1, X_2, \ldots X_n>$ into the predictive, contextual and perhaps metacontextual attributes. According to this division the levels of the Metanetwork will be built. Research in the context learning is rather active nowadays. Several fundamental works are published in this field and suggest the criteria for detecting the contextual variables, e.g. [11, 13]. We are using these criteria as they are presented in these works. We consider metacontextual variables as contextual variables for contextual variables.

Stage 2. Learning the network structure at the current level can be made by existing methods [2, 4, 5]. If the node ordering is known then the method of Cheng and Greiner is rather attractive and easy [2]. We used this method in our experiments.

It is worth to mention that for the R-Metanetwork the stage 2 returns only the maximal size model. Later, when the Metanetwork will be in use, the smaller substructure can be used according to the learned relevancies of attributes.

Stage 3. Learning the interlevel links between the current and subsequent levels. This is a new stage that has been added specifically for a Bayesian Metanetwork learning. This stage is described below both for the C-Metanetwork and for the R-Metanetwork.

Stage 4. Learning the parameters in the network at the current level is made by the standard procedure just taking into account the dynamics of parameters' values in different contexts.

If the context level is not empty, then the stages 2, 3 and 4 are repeated for the next-level network.

3.1 Learning Interlevel Links in C-Metanetwork

In Section 2, we have noticed that the vertex of every next level in a Bayesian C-Metanetwork is associated with the possible conditional probability matrix of the Bayesian network from the previous level. We will describe the establishment of such interlevel links in a C-Metanetwork.

Consider the fragment of the C-Metanetwork from Figure 3. If the standard parameter learning algorithm knows the causal relationships between the variables it will return the single conditional probability table $P(B|A)$: $P_{ik}(B = b_i \mid A = a_k)$ for the arc A→B and the single conditional probability table $P(Y \mid X)$: $P_{rs}(Y = y_r \mid X = x_s)$ for the arc X→Y. The standard algorithm processes the whole training set <*A*, *B*, *X*, *Y*>.

Assume there are several contexts in which this network fragment is observed. The parameters of the network in different contexts most probably will be different as well. In such a case it is reasonable to study each context separately and to calculate the more accurate parameters in each context instead of "averaging" probabilities over all the contexts.

Divide the whole vector <*A*, *B*, *X*, *Y*> of the training set into *n* clusters $<A, B, X, Y>_1$, $<A, B, X, Y>_2$, …,$<A, B, X, Y>_n$ according to the values of context attributes for the causal dependence $A \rightarrow B$. Applying the learning procedure in each data cluster $<A, B, X, Y>_j$ we get separately *n* conditional probability matrixes $P_j(B \mid A)$: $P_{j\,ik}(B = b_i \mid A = a_k),\ j = \overline{1,n}$.

In the same way we divide the vector <*A*, *B*, *X*, *Y*> into *m* clusters $<A, B, X, Y>_1$, $<A, B, X, Y>_2$, …,$<A, B, X, Y>_m$ according to the values of context attributes for causal dependence $X \rightarrow Y$ and get separately m conditional probability matrixes $P_t(Y \mid X)$: $P_{t\,rs}(Y = y_r \mid X = x_s)$, $t = \overline{1,m}$.

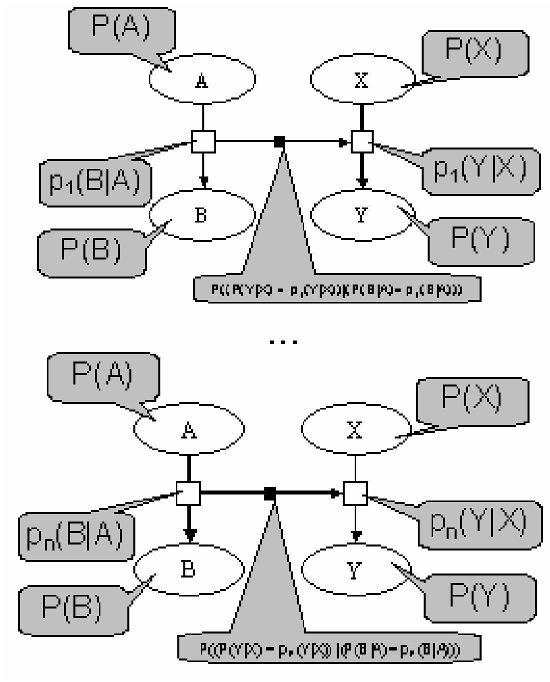

Figure 3. Different probability tables corresponding to different contexts are associated with vertexes of the second-level Bayesian network

Thus in each context the Bayesian network gets different parameters (Figure 3) and will be managed by the second-level contextual network.

The second level of the Metanetwork is entered for management of the probability tables in the first-level network. The sets of matrixes $\{P_j(B|A)\}$ and $\{P_t(Y|X)\}$ can be considered as the random variables $\boldsymbol{U}$ and $\boldsymbol{W}$ at the second level of the Metanetwork. The variable $\boldsymbol{U}$ will have as many values, as we consider contexts for the causal dependence $A \rightarrow B$. Each value u_j will correspond to the probability matrix $P_j(B|A)$. In the same way we define the variable $\boldsymbol{W}$ with the values w_t which correspond to $P_t(Y|X)$.

If the causal probabilistic dependence occurs between the contextual variables, then we learn Bayesian dependence $\boldsymbol{W} \rightarrow \boldsymbol{U}$ at the second level of the Metanetwork. The learning procedure will result in composing the (meta)matrix $P(\boldsymbol{W}|\boldsymbol{U})$, i.e. the matrix of matrixes:

$$P(W \mid U) = \{P((W = P_j(Y \mid X)) \mid (U = P_t(B \mid A))\},\ j = \overline{1,n},\ t = \overline{1,m}.$$

3.2 Learning interlevel links in R-Metanetwork

In Section 2, we have mentioned that the vertex of each next level in the Bayesian R-Metanetwork is associated with the possible relevancies of the attributes of the previous Bayesian network. We will describe the establishment of such interlevel links in the Bayesian R-Metanetwork.

Consider the fragment of the R-Metanetwork in Figure 4.

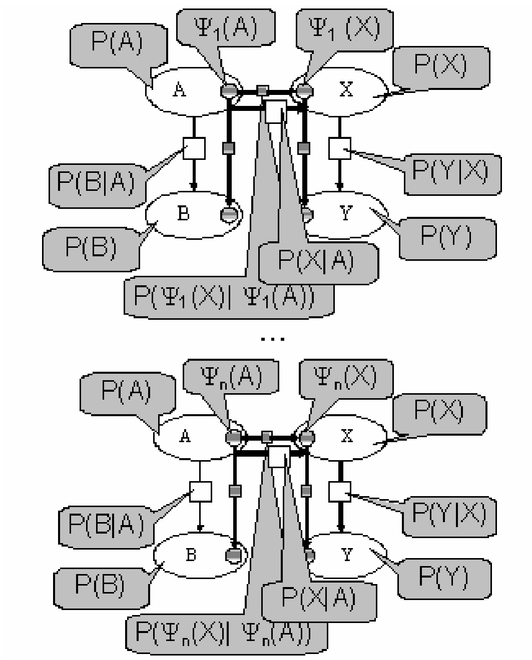

Figure 4. Different relevancies corresponding to different contexts are associated with vertexes of the second-level Bayesian network

Assume *Y* is a target attribute. The standard feature selection method will process the whole training set <*A*, *B*, *X*, *Y*> and will return some relevance estimations for each predictive attribute: $\Psi(A)$, $\Psi(B)$, $\Psi(X)$. The good overview of the existing feature selection methods is given in [3, 7]. We consider the relevance of the target attribute Ψ(Y) is equal to 1.

Assume there are several contexts in which this network fragment is observed. It is possible that relevancies of attributes will be different in different contexts. As it was done in the case of C-Metanetwork, here it is also reasonable to study each context separately and to calculate the more accurate relevancies in each context instead of "averaging" relevancies $\Psi(A)$, $\Psi(B)$, $\Psi(X)$ over all contexts. Different relevancies can lead to different network structures in different contexts.

Divide the whole vector <*A*, *B*, *X*, *Y*> of the training set into *n* clusters <*A*, *B*, *X*, *Y*>$_1$, <*A*, *B*, *X*, *Y*>$_2$, …,<*A*, *B*, *X*, *Y*>$_n$ according to the values of context attributes. Applying the learning procedure in each data cluster <*A*, *B*, *X*, *Y*>$_j$ we get separately *n* values of relevancies $\Psi_j(A)$, $\Psi_j(B)$, $\Psi_j(X)$.

Thus we get in each context the Bayesian network with different relevancies of attributes (Figure 4). The second level of the Metanetwork is entered for management of the feature selection in the first-level network. The sets of matrixes $\{\Psi_j(A)\}$, $\{\Psi_j(B)\}$, $\{\Psi_j(X)\}$ can be considered as the random variables ***U, V*** and ***W*** at the second level of the Metanetwork.

If the causal probabilistic dependence occurs between the contextual variables, then we learn the Bayesian dependence ***W→U*** at the second level of the Metanetwork. The learning procedure will result in composing the (meta)matrix $P(\boldsymbol{W}|\boldsymbol{U})$ as follows:

$$P(W \mid U) = \{P((W = \psi_j(X)) \mid (U = \psi_t(A)) = \psi_{\mathrm{j\,t}}(X \mid A)\},\ j = \overline{1,n},\ t = \overline{1,m}.$$

4. CONCLUSIONS

The general concept and the two types of a Bayesian Metanetwork are considered as tools to present the second order uncertainty. C-Metanetwork allows managing the conditional dependencies of the Bayesian network and assumes context-based conditional dependencies between conditional dependencies. R-Metanetwork assumes that the relevancies of predictive variables in the Bayesian network are random variables themselves. This metanetwork provides a tool for recalculating attributes' relevancies depending on context change. Generally a Bayesian Metanetwork is a multilevel structure of component Bayesian networks. The controlling extra level(s) in a Metanetwork is used to select the appropriate parameters or substructure from the basic network based on the contextual attributes. The accent in this paper is done to the learning procedure for Bayesian Metanetworks. Both structure learning and parameter learning are considered. The main challenge of this work is the extension of the standard Bayesian learning procedures with the algorithm of learning the interlevel links. The experiments (made outside the scope of this paper due to the domain specifics) on the data from the highly-contextual domain have shown the effectiveness of the proposed models and learning procedures. The multiple-factor concept of radiation risk for population has been modelled, and the leaning procedure has shown quite good correlation of the predicted results with expert estimations. The subjective (social) factors, which influence the radiation risk distribution, have been modelled at the contextual level of the Metanetwork. Still more experiments are needed to support the concept of a Bayesian Metanetwork and to specify concrete areas where its implementation will be reasonable. Just now we have some evidence to assume Bayesian Metanetwork to be a powerful tool in cases where structure (or strengths) of causal relationships between observed parameters of an object essentially depends on a context. Also it can be a useful diagnostic model for such an object, which diagnosis depends on different set of observed parameters depending on a context.

REFERENCES

[1] Boutiler, C., Friedman, N., Goldszmidt, M., Koller, D., Context-Specific Independence in Bayesian Networks, In: Proceedings of the 12-th Conference on Uncertainty in Artificial Intelligence (UAI-96), 1996, 115-123.

[2] Cheng J., Greiner R.. Learning Bayesian Belief Network Classifiers: Algorithms and System, In: Proceedings of the Fourteenth Canadian Conference on Artificial Intelligence (AI-2001), 2001, 141-151.

[3] Dash, M., Liu, H., Feature Selection for Classification, Intelligent Data Analysis 1 (3) 1997, 131-156.

[4] Heckerman, D., A Tutorial on Learning with Bayesian Networks. Technical Report MSR-TR-95-06, Microsoft Research, March, 1995.

[5] Madigan D., Raftery A., Volinsky C., Hoeting, J. Bayesian Model Averaging, In: Proceedings of the AAAI Workshop on Integrating Multiple Learned Models, Portland, 1996, 77-83.

[6] Pena, J., Lozano, J.A., Larranaga, P., Learning Bayesian Networks for Clustering by Means of Constructive Induction, Machine Learning, 47(1), 2002, 63-90.

[7] Siedlecki W., Skalansky J., On Automatic Feature Selection, International Journal of Pattern Recognition and Artificial Intelligence, 2, 1988, 197-220.

[8] Terziyan V., Multilevel Models for Knowledge Bases Control and Automated Information Systems Applications, Doctor of Technical Sciences Degree Thesis, Kharkov State Technical University of Radioelectronics, 1993.

[9] Terziyan V., Puuronen S., Reasoning with Multilevel Contexts in Semantic Metanetworks, In: P. Bonzon, M. Cavalcanti, R. Nossun (Eds.), Formal Aspects in Context, Kluwer Academic Publishers, 2000, 97-126.

[10] Terziyan V., Vitko O., Bayesian Metanetworks for Modelling User Preferences in Mobile Environment. In: Proceedings of the German Conference on Artificial Intelligence (KI-2003), Hamburg, 15-18 September 2003 , LNCS, Vol. 2821, Springer, 370-384.

[11] Turney P. The identification of context-sensitive features: a formal definition of context for concept learning, In: Proceedings of the Workshop on Learning in Context-Sensitive Domains at the Thirteenth International Conference on Machine Learning (ICML-96),, 1996, 53-59.

[12] Vitko O., The Multilevel Probabilistic Networks for Modelling Complex Information Systems under Uncertainty. Ph.D. Thesis, Kharkov National University of Radioelectronics, 2003.

[13] Widmer G. Tracking Context Changes through Meta-Learning, Machine Learning, 27(3), 1997, 259-286.