Directed Cycles in Bayesian Belief Networks: Probabilistic Semantics and Consistency Checking Complexity

Alexander L. Tulupyev and Sergey I. Nikolenko

St.Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences
alt@iias.spb.su
sergey@logic.pdmi.ras.ru

Abstract. Although undirected cycles in directed graphs of Bayesian belief networks have been thoroughly studied, little attention has so far been given to a systematic analysis of directed (feedback) cycles. In this paper we propose a way of looking at those cycles; namely, we suggest that a feedback cycle represents a family of probabilistic distributions rather than a single distribution (as a regular Bayesian belief network does). A non-empty family of distributions can be explicitly represented by an ideal of conjunctions with interval estimates on the probabilities of its elements. This ideal can serve as a probabilistic model of an experts uncertain knowledge pattern; such models are studied in the theory of algebraic Bayesian networks. The family of probabilistic distributions may also be empty; in this case, the probabilistic assignment over cycle nodes is inconsistent. We propose a simple way of explicating the probabilistic relationships an isolated directed cycle contains, give an algorithm (based on linear programming) of its consistency checking, and establish a lower bound of the complexity of this checking.

1 Introduction

Bayesian belief networks (BBN), originating in the works of Judea Pearl [19,20,21], have employed directed acyclic graphs (DAG) in order to describe a probabilistic distribution in a way convenient for bayesian inference (we refer to [13] for an excellent overview of the subject). Since the original works, many generalizations and similar apparata have been developed, among them being, for example, dynamic Bayesian networks (see [14] and references therein).

Many efforts went into generalizing the basic structure of the network. Always the generalizations were related to employing a more general structure to be able to build more general independency models, that is, to incorporate different statements of the kind "X is independent of Y given Z". Finally, chain graphs as described in [23] seem to solve this problem (although there is still plenty of room for improvement). They have a complex structure with three different kinds of edges and allow undirected cycles in the graph.

A. Gelbukh, A. de Albornoz, and H. Terashima (Eds.): MICAI 2005, LNAI 3789, pp. 214–223, 2005.

Returning to directed acyclic graphs, the current state of the art in the Bayesian belief networks allows to efficiently deal with undirected cycles, that is, patterns which would be cycles if the arrow directions were not taken into account. However, very little (if any) work seems to have been done in the direction of generalizing Bayesian belief networks to allow directed cycles. The article [22], despite its highly relevant title, deals with establishing Markov properties of directed cyclic graphs representing stochastically disturbed linear equations, and does not deal with semantics of a cycle in a Bayesian belief network. It is also clearly stated in [13] that there is no BBN-calculus developed to deal with directed (feedback) cycles in Bayesian belief networks.

However, this generalization seems natural and may occur in practice, since the structure of a Bayesian network is determined by the experts. In fact, the need for this generalization has already been encountered in literature: see, for example, [1]. In that article the authors simply revert an edge of the directed cycle; however, such an operation, as we shall show here, changes the semantics of the whole network and is certainly not the right thing to do.

In this paper, we consider semantics of a directed cycle in a Bayesian belief network; the network is defined over a set of atomic propositions. We show that a cycle introduces interval bounds for the joint probabilities of variables of the network, and thus requires a new formalism to deal with it. We shall need to deal with a whole family of distributions, which may be empty. Thus, we shall look for algorithms that check consistency and find the upper and lower bounds for marginal joint probabilities of the cycle's elements.

Our approach, in a certain sense, is a complement to Heckerman *et al.* [11,12]. They are ready to work with as many cycles as may appear in their *dependency networks* that may also be, as they say, *almost consistent* (instead of being just consistent with probability axioms). We are incorporating cycles in BBN calculi preserving consistency with probabilistic axioms and avoid using artificial constraints requiring strict positiveness of appearing probabilistic distributions, as opposed to Heckerman *et al.*

2 Basic Definitions

In the paper, we follow a probabilistic logic approach introduced by N. Nilsson in [18] and formalized from a logical viewpoint in [2,3,4].

Let $\mathcal{T} = \{t_1, \ldots, t_N\}$ be the set of atoms (atomic propositions, Boolean variables) that are to represent experts' elementary judgements about a certain domain. $\mathcal{S} = \{x_1, \ldots, x_n\}$ is a subset of $\mathcal{T}$: $\mathcal{S} \subseteq \mathcal{T}$. We denote the negation of x by $\bar{x}$.

An ideal of conjunctions $\mathcal{C} = \mathcal{C}(\mathcal{S})$ over $\mathcal{S}$ consists of all non-empty conjunctions of elements of $\mathcal{S}$. For example,

$$\mathcal{C}(\{x_1, x_2, x_3\}) = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1x_2x_3\}.$$

A consistent probabilistic distribution over $\mathcal{C}(\mathcal{S})$ can be uniquely extended to all propositional formulas built over $\mathcal{S}$. A consistent assignment of point-valued estimates of probabilities of the elements of $\mathcal{C}$ defines the unique probabilistic distribution over propositional formulas over $\mathcal{S}$. A consistent assignment of interval-valued estimates defines a family of probabilistic distributions. The algorithms for consistency checking for such a distribution (or a family of the distributions) have already been developed, and we shall use them to cope with a cycle in a BBN. Let us note in addition that ideal $\mathcal{C}$ can be considered as a probabilistic model for an expert's knowledge pattern with uncertainty.

A *Bayesian belief network* is traditionally defined as a directed acyclic graph (DAG) $G = (V, E)$ (where V is a finite set of nodes, and E is a set of edges, that is, $E \subseteq G \times G$) together with a joint probabilistic distribution P that satisfies the *Markov condition*, namely that each variable $x \in V$ is conditionally independent of the set of all its nondescendents given the set of all its parents (see [15] for this definition and a detailed consideration of BBNs). The probabilistic distribution in question is defined by assigning conditional probabilities to each node given its parents.

In brief, an *algebraic Bayesian network* (ABN) is a set of possibly intersecting ideals of conjunctions together with point-valued or interval-valued estimates on the joint probabilities of conjunctions appearing in these ideals. Formally, a BBN's knowledge pattern is modeled with a point-valued tensor of conditional probability. In contrast to BBN, an ABN's knowledge pattern is modeled with an ideal of conjunctions that represents marginal probabilities in a specific form. Those marginal joint probabilities may be assigned with point-valued or interval-valued estimates.

In what follows we are trying to extend the class of BBNs with directed cyclic graphs, keeping the method of defining a network. It turns out that together with the word "acyclic" we shall need to throw away the concept of having a *single* probabilistic distribution corresponding to a BBN. We should rather consider families of distributions corresponding to a cycle in a BBN.

3 Semantics of a Cycle

In this article we restrict ourselves to the simplest cyclic situation possible: a generalized Bayesian belief network consisting of a single directed cycle. We denote the nodes of the graph by $x_1, \ldots, x_n$. The cycle is presented on Fig. 1.

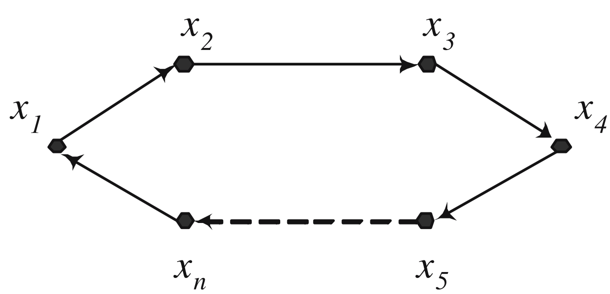

Fig. 1. An isolated cycle with n vertices

By definition, we receive as input the probabilities

$$p(x_1 \mid \tilde{x}_n), p(x_2 \mid \tilde{x}_1), \ldots, p(x_n \mid \tilde{x}_{n-1}), \quad \tilde{x}_i \in \{x_i, \bar{x}_i\},\ i \in 1(1)n.$$

Note that there are no additional restrictions by the Markov condition in this case (because the set of nondescendents of each node in an isolated directed cycle is empty).

Let us try to deduce from the given data the marginal probabilities and the description of the probabilistic distribution this network represents. By the definition of conditional probability we can obtain the following equations:

$$\begin{cases} p(x_1) = p(x_1 \mid x_n)p(x_n) + p(x_1 \mid \bar{x}_n)(1 - p(x_n)) \\ p(x_2) = p(x_2 \mid x_1)p(x_1) + p(x_2 \mid \bar{x}_1)(1 - p(x_1)) \\ \quad\vdots \\ p(x_n) = p(x_n \mid x_{n-1})p(x_{n-1}) + p(x_n \mid \bar{x}_{n-1})(1 - p(x_{n-1})) \end{cases} \tag{1}$$

Note that the only unknowns in this system are the probabilities $p(x_i)$. Thus, the system is a linear system (with a very simple structure) of the form

$$\begin{pmatrix} 1 & 0 & 0 \ldots & 0 & r_{1n} \\ r_{21} & 1 & 0 \ldots & 0 & 0 \\ 0 & r_{32} & 1 \ldots & 0 & 0 \\ \vdots & \vdots & \vdots \ddots & \vdots & \vdots \\ 0 & 0 & 0 \ldots & r_{n,n-1} & 1 \end{pmatrix} \begin{pmatrix} p(x_1) \\ p(x_2) \\ p(x_3) \\ \vdots \\ p(x_n) \end{pmatrix} = \begin{pmatrix} p(x_1 \mid \bar{x}_n) \\ p(x_2 \mid \bar{x}_1) \\ p(x_3 \mid \bar{x}_2) \\ \vdots \\ p(x_n \mid \bar{x}_{n-1}) \end{pmatrix}, \tag{2}$$

where $r_{ij} = p(x_i \mid \bar{x}_j) - p(x_i \mid x_j)$. Thus, we obtain $p(x_i)$.

Remark 1. There is a special case when the system is degenerate. This may happen if the determinant of the matrix of the system is equal to zero, that is, if $1 - r_{1n}r_{21} \ldots r_{n,n-1} = 0$. This is possible only if all $r_{i,i-1} = \pm 1$. If in this case the right-hand side is non-zero, the system has no solutions, and the network is inconsistent. If the right-hand side is zero, it means that *de facto* all nodes of the network describe the same judgement x. There are no restrictions on the probability of that x, that is, all we can say is $p(x) \in [0, 1]$. In this case, of course, there is no need to draw a network, let alone a directed cycle, so we may assume that this case does not hold in practice (if it happens, the cycle may be easily reduced to one node without loss of information).

After we have calculated $p(x_i)$, we may proceed to find $p(x_i x_{i-1})$, $i = 1, \ldots, n$, by definition of conditional probability (we denote here $x_0 := x_n$ for simplicity of the formulae and freely substitute x_0 and x_n further):

$$p(x_i x_{i-1}) = p(x_{i-1})p(x_i \mid x_{i-1}) + (1 - p(x_{i-1}))p(x_i \mid \bar{x}_{i-1}). \tag{3}$$

But this is exactly where our certain knowledge about the point-valued probabilities stops.

Joint probabilities of three and more variables, and even joint probabilities of pairs of variables representing non-adjacent graph nodes may *not* in general be determined from the input. Only interval bounds may be established as a result of solving the linear programming task that will be described in Sect. 5.

4 Special Case: A Cycle with Two Vertices

The simplest (and in some ways special) case of a cyclic BBN is an isolated cycle with two vertices. This case is special because it is the only case where the input conditional probabilities determine the whole probabilistic distribution uniquely. The marginal probabilities $p(x_1)$ and $p(x_2)$ satisfy the following linear system (a special case of (2)):

$$\begin{cases} p(x_1) + (p(x_1|x_2) - p(x_1|\bar{x}_2))p(x_2) = p(x_1|\bar{x}_2) \\ (p(x_2|x_1) - p(x_2|\bar{x}_1))p(x_1) + p(x_2) = p(x_2|\bar{x}_1) \end{cases} \tag{4}$$

After solving it we obtain explicit formulae for marginal probabilities:

$$p(x_1) = \frac{p(x_1|\bar{x}_2) - (p(x_1|x_2) - p(x_1|\bar{x}_2))p(x_2|\bar{x}_1)}{1 - (p(x_1|x_2) - p(x_1|\bar{x}_2))(p(x_2|x_1) - p(x_2|\bar{x}_1))}$$

$$p(x_2) = \frac{p(x_2|\bar{x}_1) - (p(x_2|x_1) - p(x_2|\bar{x}_1))p(x_1|\bar{x}_2)}{1 - (p(x_1|x_2) - p(x_1|\bar{x}_2))(p(x_2|x_1) - p(x_2|\bar{x}_1))}$$

It is now easy to obtain the entire distribution:

$$\begin{aligned} p(x_1x_2) &= p(x_1)p(x_2|x_1), \\ p(x_1\bar{x}_2) &= p(x_1)(1 - p(x_2|x_1)), \\ p(\bar{x}_1x_2) &= (1 - p(x_1))p(x_2|\bar{x}_1), \\ p(\bar{x}_1\bar{x}_2) &= (1 - p(x_1))(1 - p(x_2|\bar{x}_1)). \end{aligned} \tag{5}$$

Let us note that these formulae make sense provided

$$(p(x_1|x_2) - p(x_1|\bar{x}_2))(p(x_2|x_1) - p(x_2|\bar{x}_1)) \neq 1,$$

which may happen only in the degenerate case considered in the previous section.

5 Consistency Checking

As we have seen above, a Bayesian belief network with a directed cycle may describe a whole family of probabilistic distributions rather than the only one, as a regular Bayesian belief network does. This family may, of course, be empty. Therefore, the problem of establishing the consistency of an initial probabilistic assignment arises. In this section we describe the most straightforward way to check for consistency. However, as we show further, this method is hard to improve. In this section, as in the whole article, we restrict ourselves to the case of one isolated cycle on n nodes.

To establish consistency, we must ensure that there exists a probabilistic distribution $p(\tilde{x}_1\tilde{x}_2\ldots\tilde{x}_n)$ over all n variables which is compatible with axioms of probability

$$\forall \tilde{x}_1\tilde{x}_2\ldots\tilde{x}_n \quad p(\tilde{x}_1\tilde{x}_2\ldots\tilde{x}_n) \geq 0; \qquad \sum_{\tilde{x}_1\tilde{x}_2\ldots\tilde{x}_n} p(\tilde{x}_1\tilde{x}_2\ldots\tilde{x}_n) = 1$$

and meets conditional probabilities given as input. To do that, we simply solve the linear programming task that may be extracted from the axioms of probability and given constraints. Its unknowns are probabilities of positive conjunctions $p(x_{i_1}x_{i_2}\dots x_{i_k})$. The input enters the formulation of the linear programming task as $p(x_i)$ and $p(x_ix_j)$, which, as was shown above, may be deduced from the input.

Example 1. We provide the linear programming task for the case of a cycle on three nodes. In this case the problem is trivial because there is only one variable, $p(x_1x_2x_3)$. We should minimize and maximize it over the following set of constraints:

$$\begin{cases} p(x_1x_2x_3) \geq 0 \\ p(x_1x_2) - p(x_1x_2x_3) \geq 0 \\ p(x_1x_3) - p(x_1x_2x_3) \geq 0 \\ p(x_2x_3) - p(x_1x_2x_3) \geq 0 \\ p(x_1) - p(x_1x_2) - p(x_1x_3) + p(x_1x_2x_3) \geq 0 \\ p(x_2) - p(x_1x_2) - p(x_2x_3) + p(x_1x_2x_3) \geq 0 \\ p(x_3) - p(x_1x_3) - p(x_2x_3) + p(x_1x_2x_3) \geq 0 \\ 1 - p(x_1) - p(x_2) - p(x_3) + p(x_1x_2) + p(x_2x_3) + p(x_1x_3) - p(x_1x_2x_3) \geq 0 \end{cases}$$

By solving similar linear programming problems, we may establish interval bounds for the probabilities of unknown conjunctions in the general case. These will allow us to reconstruct the overall probability distribution (that is, the family of distributions).

Example 2. We show how conjunctions of positive literals generate the entire distribution for the case of three variables:

$$\begin{cases} p(\bar{x}_1x_2x_3) = p(x_2x_3) - p(x_1x_2x_3) \\ p(x_1\bar{x}_2x_3) = p(x_1x_3) - p(x_1x_2x_3) \\ p(x_1x_2\bar{x}_3) = p(x_1x_2) - p(x_1x_2x_3) \\ p(\bar{x}_1\bar{x}_2x_3) = p(x_3) - p(x_2x_3) - p(x_1x_3) + p(x_1x_2x_3) \\ p(\bar{x}_1x_2\bar{x}_3) = p(x_2) - p(x_1x_2) - p(x_2x_3) + p(x_1x_2x_3) \\ p(x_1\bar{x}_2\bar{x}_3) = p(x_1) - p(x_1x_2) - p(x_1x_3) + p(x_1x_2x_3) \\ p(\bar{x}_1\bar{x}_2\bar{x}_3) = 1 - p(x_1) - p(x_2) - p(x_3) + \\ \qquad + p(x_1x_2) + p(x_2x_3) + p(x_1x_3) - p(x_1x_2x_3). \end{cases}$$

However, the linear programming problem is in general very large. For a cycle of n nodes, it has 2^n constraints and $2^n - 2n - 1$ unknowns ($2n$ unknowns disappear since we can determine $p(x_i)$ and $p(x_ix_{i-1})$). Solving it for large knowledge patterns would require too much computational power. It would be extremely helpful to reduce this task to some easier ones. However, in the next section we give a negative result on this approach.

Before we proceed to complexity issues, we should remark on the nature of the result. What we had to do in order to check consistency and establish the interval bounds, is known as the *a priori inference* in the theory of algebraic Bayesian networks introduced by V. Gorodetski in [5,6,7,8,9,16] and developed in [10,24,25]. We simply immersed the cycle in question into the corresponding knowledge pattern of an algebraic Bayesian network. It is not the point of this article to compare the two formalisms, but in this case algebraic Bayesian networks turn out to be more descriptive than Bayesian belief networks, because they are able to capture this kind of relation between the boolean variables.

6 Complexity of the Consistency Checking

As we have seen in the previous section, consistency checking in general is an expensive task. Since Bayesian networks were intended to deal with decomposable distributions, a natural question arises: are we able to decompose the problem on the big knowledge pattern that includes the whole cycle to smaller problems on some subsets of the cycle?

The answer is definitely negative. Moreover, we note that, in fact, consistency of a cycle *always* has to be considered as a whole, rather than in part. The matter is that a linear chain of nodes in a Bayesian belief network is always consistent (see [13] or any other source on Bayesian belief networks); it usually comes with a number of conditional independence restrictions that allow to single one particular distribution out of the whole family, but the family is never empty anyway.

However, as soon as we engage cycles (even isolated), inconsistent cycles become possible. We give here an example of an inconsistent cycle on three vertices.

Example 3. Consider the following Bayesian belief network — a cycle on three vertices:

$$p(x_2|x_1) = 1/4, p(x_3|x_2) = 3/4, p(x_1|x_3) = 3/4,$$
$$p(x_2|\bar{x}_1) = 1/2, p(x_3|\bar{x}_2) = 1/6, p(x_1|\bar{x}_3) = 1/6.$$

By solving the linear system (2), we obtain

$$\begin{aligned} &p(x_1) = p(x_2) = p(x_3) = \tfrac{2}{5}, \\ &p(x_1x_2) = \tfrac{1}{10}, \\ &p(x_2x_3) = p(x_1x_3) = \tfrac{3}{10}. \end{aligned}$$

Now restrictions on $p(x_1x_2x_3)$ include, on one hand, $p(x_1x_2) - p(x_1x_2x_3) \geq 0$, that is, $p(x_1x_2x_3) \leq 1/10$, and, on the other hand, $p(x_3) - p(x_1x_3) - p(x_2x_3) + p(x_1x_2x_3) \geq 0$, that is, $p(x_1x_2x_3) \geq 2/10$. Thus, this cycle on three vertices is inconsistent.

Therefore, even if more efficient algorithms exist (which they may), they should take into consideration all the data at once, and consider the overall probabilistic distribution.

7 On Reverting Edges in a Cycle

A cycle, as we have shown above, forces the probabilistic semantics of a Bayesian belief network outside the realm of unique distributions that seems so natural for

Bayesian belief networks. Therefore, it is natural that the basic idea of previously suggested ways to cope with cycles has been to try and get rid of the cycle and thus reduce the problem to well-known cases. We have already mentioned in the introduction the paper [1], where the simplest way to remove a cycle is considered. The proposed technique is to revert an edge in the cycle, making it non-directed and, therefore, subject to standard Bayesian network analysis.

However, this is incorrect, because by reverting an edge (and considering the result as a regular Bayesian belief network) we would replace a whole family of distributions by a single one. In fact, the initial network might be inconsistent, but the result will always be consistent. The distribution becomes unique because reverting an edge imposes additional constraints in the form of conditional independence of certain nodes of the cycle that are now (after reverting an edge) d-separable. However, even if the initial cycle was consistent, the unique distribution appearing after reverting an edge might not even be contained in the initial family of distributions — at least, it has to be proven. There is no sound justification for this process in [1].

Such a justification might be that this unique distribution has some special properties which single it out of the family. In [17] we consider ways to look for this distribution. For example, it may be a good idea to select distributions based on the maximal entropy principle. There also exist experimental techniques for selecting a single distribution, for example, stochastic modeling. However, this problem remains open — even in the motivational phase, it is not clear what kind of a distribution to look for (since the usual decomposable ones don't work anymore).

One certainly valid way to work with a cycle is to revert all its edges at once (leaving the cycle in place, but changing its direction). This does not change the semantics of the network, because receiving as input conditional probabilities $\{p(\tilde{x}_i|\tilde{x}_{i-1})\}$, as we have shown above, to receiving joint probabilities $\{p(\tilde{x}_i), p(\tilde{x}_i\tilde{x}_{i-1})\}$, and the latter do not depend on where the edges in the cycle are directed.

For an isolated cycle this reverting is, of course, meaningless. However, it may prove useful for coping with several intersecting cycles or other Bayesian belief networks with more complex structure — for example, it may simplify this structure, remove unnecessary cycles, and so on.

8 Conclusions and Further Work

We have shown that a cycle in a BBN may represent a (possibly empty) family of probabilistic distributions over its elements and their conjunctions, rather than a single distribution a regular BBN represents. If we want to incorporate cycles into a BBN, we should therefore check the network for consistency. In this article this work is done for the case of an isolated cycle. We have also shown that to establish consistency in general it is necessary to consider the exponential-sized linear programming task, since no smaller one (none corresponding to the joint distribution of a smaller set of variables) would suffice.

All known BBN calculi deal with a single distribution defined by the network. Therefore, we should either somehow choose a single distribution out of the family of possible distributions, or generalize the calculus to deal with interval-valued estimates of probabilities of BBN nodes, as is done in the ABN approach.

One of the natural directions for further work would be to generalize the algorithms of evidence propagation to the directed cyclic case. This approach has been to some extent carried out in the theory of algebraic Bayesian networks, but there are differences between the two formalisms that make considering cyclic Bayesian belief networks a worthwhile task by itself.

However, Bayesian belief networks on directed cyclic graphs should always be treated with care, that is, one should carefully check for consistency and deal with families of distributions rather than a single one. This feature will not disappear unless too restrictive conditions are satisfied (and thus the formalism is rendered impractical). Therefore, another direction for future research is to try to establish sufficient conditions for effective consistency checking. In the current work we have shown that no effective necessary conditions exist, but some good enough sufficient conditions may cover many interesting cases.

References

1. Castillo, E., Gutierrez, J. M., Hadi, A. S. Modeling Probabilistic Networks of Discrete and Continuous Variables. *Journal of Multivariate Analysis*, **64** (1998), pp.48–65.
2. Fagin, R., Halpern, J. Y., Megiddo, N. A Logic for Reasoning about Probabilities. Report RJ 6190(60900) 4/12/88, pp. 1–41.
3. Fagin R., Halpern J. Y. Uncertainty, Belief, and Probability. Proc. of the 11th International Joint Conference on Artificial Intelligence (1989), pp. 1161–1167.
4. Fagin R., Halpern J. Y. Uncertainty, Belief, and Probability-2. Proc. of the IEEE Simposium on Logic and Computer Science, **7** (1991), pp.160–173.
5. Gorodetski, V. Adaptation Problems in Expert System. *International Journal of Adaptive Control and Signal Processing*, **6** (1992), pp.201–210.
6. Gorodetski, V. Expert Systems Adaptation. *Transactions of the Russian Academy of Sciences "Technical Cybernetics"*, **5** (1993), pp.101–110 (in Russian).
7. Gorodetski, V. Interval-valued Probabilistic Measure of Uncertainty in Knowledge Engineering. In Yusupov R.M. (ed.), *Theoretical Basis and Applied Intelligent Information Technologies*. Russian Academy of Science, SPIIRAS (1998), pp.44–58 (in Russian).
8. Gorodetski, V. I., Nesterov, V. M. Interval probabilities and knowledge engineering. In: Alefeld, G., Trejo, R. A. (eds), *Interval computations and its applications to reasoning under uncertainty, knowledge representation and control theory*, Mexico City (1998), pp.15–20.
9. Gorodetski, V., Skormin, V., Popyack, L. Data Mining Technology for Failure Prognostics of Avionics. *IEEE Transactions on Aerospace and Electronic Systems*, **38**, no. 2 (2002), pp.388–403.
10. Gorodetskii, V. I., Tulupyev [Tulup'ev], A. L. Generating Consistent Knowledge Bases with Uncertainty. *Journal of Computer and Systems Sciences International*, **5**, 1997, pp.33–42.

11. Heckerman, D., D. Chickering, D. M., Meek, C., Rounthwaite, R., Kadie, C. Dependency Networks for Density Estimation, Collaborative Filtering, and Data Visualization. *Journal of Machine Learning Research*, **1**:49–75, 2000. Also appears as Technical Report MSR-TR-00-16, Microsoft Research, February, 2000.
12. Hulten, G., Chickering, D. M., Heckerman, D. Learning Bayesian Networks from Dependency Networks: A Preliminary Study. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, Key West, FL, 2003.
13. Jensen, F. V. Bayesian networks and decision graphs. Springer-Verlag, 2002.
14. Mihajlovic, V., Petkovic, M. Dynamic Bayesian Networks: a state of the art. Technical report, 2001, no. TR-CTIT-01-34, Centre for Telematics and Information Technology, University of Twente.
15. Neapolitan, R. E. Learning Bayesian Networks. Pearson Prentice Hall, 2004.
16. Nesterov, V. M., Gorodetski, V. I. Interval algorithm for checking the consistency of the algebraic Bayesian net. In: Alefeld, G., Trejo, R. A. (eds), *Interval computations and its applications to reasoning under uncertainty, knowledge representation and control theory*. Mexico City (1998), pp.28–29.
17. Nikolenko, S. I., Tulupyev, A. L. Edge flipping as a technique for dealing with directed cycles in Bayesian networks. *MEPhI-2005 Science Session Proceedings*, vol. 3, Intellectual systems and technologies, pp. 176–178. Moscow, MEPhI, 2005 (in Russian).
18. Nilsson, N.J. Probabilistic Logic. *Artificial Intelligence*, **28** (1986), pp. 71–87.
19. Pearl, J. How to do with probabilities what people say you can't. In S. Quaglini, P. Barahona, S. Andreassen (eds.) *Artificial Intelligence in Medicine*, LNAI2101 (2001), 283–292.
20. Pearl, J. Probabilistic reasoning using graphs. In B. Bouchon, R.R. Yager (eds.) *Uncertainty in Knowledge-Based Systems*, Springer-Verlag (1987), 201–202.
21. Pearl, J. Probabilistic reasoning in intelligent systems: networks of plausible inference. San Mateo, Morgan Kauffmann, 1988.
22. Sprites, P. Directed cyclic graphical representations of feedback models. *Uncertainty in Artificial Intelligence*, **12** (1996), pp.454–461.
23. Studený, M., Bouckaert, R. On chain graph models for description of conditional independence structures. *The Annals of Statistics*, **26**, No. 4 (1998), 1434–1495.
24. Tulupyev, A.L. Algebraic Bayesian Networks: Theoretical Background and Consistency. St.-Petersburg, 1995 (in Russian).
25. Tulupyev, A. L. Algebraic Bayesian Networks: a Probabilistic Logic Approach to Modeling Knowledge Bases with Uncertainties. St.-Petersburg, 2000 (in Russian).

ERRATUM

to

Tulupyev A.L., Nikolenko S.I. Directed Cycles in Bayesian Belief Networks: Probabilistic Semantics and Consistency Checking Complexity // MICAI 2005: Advances in Artificial Intelligence**.** 4th Mexican International Conference on Artificial Intelligence, Monterrey, Mexico, November 14-18, 2005, Proceedings Series: Lecture Notes in Computer Science; Subseries: Lecture Notes in Artificial Intelligence, Vol. 3789. Gelbukh, Alexander; Terashima, Hugo (Eds.) 2005, XXVI. P. 214–223.

The formulae (3) on page 217 reads

$$p(x_{i-1}x_i) = p(x_{i-1})\,p(x_i \mid x_{i-1}) + (1 - p(x_{i-1}))\,p(x_i \mid \bar{x}_{i-1})\,. \tag{3}$$

It should read

$$p(x_{i-1}x_i) = p(x_{i-1})\,p(x_i \mid x_{i-1}). \tag{3}$$

Both authors bring their apologies for the error detection happened too late.

This *Erratum* is written by the authors and is not a part of the original publication.

Alexander L. Tulupyev
Sergey I. Nikolenko

St. Petersburg
November 26, 2005