Appl. Math. and Comp. Sci., 1996, vol.6, No.3, 529–564

AUTOMATED CONSTRUCTION OF POSSIBILISTIC NETWORKS FROM DATA[†]

JÖRG GEBHARDT*, RUDOLF KRUSE*

Possibilistic networks constitute a promising framework for efficient treatment of uncertain and imprecise information in knowledge-based systems. In this paper, we propose a new method for induction of the structure (the qualitative part) and the attached possibility distributions (the quantitative part) of a possibilistic network from a database of sample cases that may contain imprecise or missing values. It turns out that a modified random-set approach to the semantics of possibility distributions is adequate to provide a possibilistic interpretation of the databases under consideration. Since constructing a possibilistic network can be viewed as a generalization of the structure identification problem in relational data, we have to overcome well-known complexity problems. Therefore we present an efficient Greedy search structure induction algorithm for possibilistic networks that has successfully been applied to construct a non-trivial network of practical interest from a given database.

1. Introduction

One of the major problems in handling imprecise and uncertain information in knowledge-based systems is the problem of finding a computationally appealing description of the available data, which is both economical in using storage and supports efficient propagation techniques. The existence of such a description strongly depends on whether dependencies among the data items are decomposable into local, more basic dependencies. The relational database model (Maier, 1983; Ullman, 1988; 1989) has tackled this problem by storing the database as a lossless join decomposition, namely a collection of projections, from which the original relation can be reconstructed. Recent work in this field concerns structure identification in relational data (Dechter and Pearl, 1992) and the clarification of cross-references to the solution of constraint-satisfaction problems (Dechter and Pearl, 1988; Gyssens *et al.*, 1994). Although various applications in artificial intelligence, database theory, graph theory, and operational research are connected with decomposition problems, very few general results have been established so far. The most advanced deliverables refer to the induction of Bayesian networks (Lauritzen and Spiegelhalter, 1988; Pearl,

[†] This work has been partially supported by CEC-ESPRIT III Basic Research Project 6156 (DRUMS II).

* Dept. of Mathematics and Computer Science, University of Braunschweig, Bültenweg 74–75, D-38106 Braunschweig, Germany, e-mail: gebhardt@ibr.cs.tu-bs.de

1988) from statistical data. For an overview, see (Buntine, 1994). The corresponding algorithms are based on linearity and normality assumptions (Pearl and Wermuth, 1993), the extensive testing of conditional independence relations (Spirtes and Glymour, 1991; Verma and Pearl, 1992) or Bayesian approaches (Cooper and Herskovits, 1992; Lauritzen *et al.*, 1993). Some crucial problems regarding these methods are their computational complexity, their limited reliability unless the amount of data is enormous, and strong presuppositions like the requirement of an ordering of the nodes and *a priori* distribution assumptions. In order to overcome the complexity problems, some heuristic approaches like, for example, the K2 Greedy algorithm (Cooper and Herskovits, 1992) and the combination of conditional independence tests with Bayesian learning (Singh and Valtorta, 1993) have been proposed. But even finding a most likely belief network structure B_S, given a database $\mathcal{D}$ of samples, often turns out to be not sufficiently informative if the probability $P(B_S \mid \mathcal{D})$ is too small.

Bayesian networks provide a well-founded normative framework for knowledge representation and belief updating in the presence of *uncertainty* in *precise* data, but extending pure probabilistic settings to the treatment of *imprecise* (set-valued) information in general restricts the computational tractability of inference mechanisms. Therefore it seems to be convenient to discuss appropriate strategies of *information compression* and to investigate other uncertainty calculi in order to support efficient propagation without essentially affecting the expressive power and correctness of decision making procedures. Information compression is acceptable for systems that perform *approximate reasoning*, justified by a weak sensitivity with respect to an imperfect modelling of the system's behaviour.

A quite general approach to uncertain reasoning under imprecision in the so-called *valuation-based networks* which can be applied, for example, to *upper and lower probabilities* (Kyburg, 1987; Walley, 1991; Walley and Fine, 1982), *Dempster–Shafer theory of evidence* (Dempster, 1967; 1968; Shafer, 1976; Shafer and Pearl, 1990; Smets and Kennes, 1994), and *possibility theory* (Dubois and Prade, 1988; 1991; Zadeh, 1978) has been proposed in (Shenoy, 1989; 1993; Shenoy and Shafer, 1990) and implemented in the software tool PULCINELLA (Saffiotti and Umkehrer, 1991). Nevertheless, quantitative learning of distribution functions as well as qualitative learning of dependency structures in the non-standard calculi of uncertainty modelling has not been studied in much detail yet.

In this paper, we focus our attention on the possibility theory as a promising framework for the information-compressed representation of databases of (imprecise) sample cases. Compared to the alternative uncertainty calculi that we have mentioned above, it is the simplest one in order to support the development of efficient inference and learning algorithms.

As an example for the addressed type of imperfect information consider the database given in Table 1. It shows five *variables (attributes)* X_1, X_2, X_3, X_4, and X_5 that describe the state of some system. X_1, X_2, and X_4 denote three possible faults, whereas X_3 and X_5 stand for two possible findings that are influenced by one or more of the faults. The *domains* $\Omega^{(i)} = \mathrm{Dom}(X_i)$, $i = 1, \ldots, 5$, of the variables are defined as $\mathrm{Dom}(X_i) = \{0, 1\}$, where '0' reflects the absence, and '1' the presence of the corresponding item.

Tab. 1. A database $\mathcal{D}$ of cases.

case	X_1	X_2	X_3	X_4	X_5
1	1	0	1	1	1
2	0	1	0	1	$\{0,1\}$
3	0	1	1	$*$	0
4	0	1	0	0	1
5	1	0	1	1	0
6	0	1	1	0	1
7	0	1	1	1	1
8	0	1	0	1	0

Our database consists of eight representative sample cases that specify the previous states of the system, and each of them can be formalized as a tuple $\omega_j \in \Omega$, taken from the *universe of discourse* $\Omega = \Omega^{(1)} \times \ldots \times \Omega^{(5)}$.

For example, the first case of the database is a *precise* specification of the system state ω_1, stating that $X_1 = 1$, $X_2 = 0$, $X_3 = 1$, $X_4 = 1$, and $X_5 = 1$. The second case is an *imprecise* (set-valued) specification of ω_2, given by $X_1 = 0$, $X_2 = 1$, $X_3 = 0$, $X_4 = 1$, and $X_5 \in \{0,1\}$. Imprecision consists in the existence of at least two alternatives that are both considered to be possible values of X_5, while ignoring the information contained in the rest of the database. The third case is a *missing value* specification of ω_3, namely $X_1 = 0$, $X_2 = 1$, $X_3 = 1$, $X_4 = *$, and $X_5 = 0$, where '$*$' stands for an unobserved value, which may be assigned based on the dependencies that follow from the other cases.

Our aim is to use the database as general knowledge for an imperfect specification of a system state $\omega_0 \in \Omega$ that may be observed at a selected point in time. We call ω_0 the "current object state of interest." The imperfection in using the database for a specification of ω_0 is due to *uncertainty* about the selection of sample cases $R \subseteq \Omega$ that are correct specifications of ω_0 (i.e. $\omega_0 \in R$) and the *nonspecificity* of imprecise cases (i.e. $|R| > 1$). In the possibilistic setting, imperfection is described by *possibility distributions* $\pi : \Omega \to [0,1]$, where $\pi(\omega)$ quantifies the *degree of possibility* with which the statement "$\omega = \omega_0$" is regarded as being true.

Recent years of research have resulted in various proposals for the semantic background of a degree of possibility and possibility theory as a framework for reasoning with uncertain and imprecise data (Dubois *et al.*, 1993). From the viewpoint of providing an adequate possibilistic interpretation of a database of (imprecise) sample cases, it seems to be most convenient to introduce possibility distributions as *one-point coverages* of random sets (Hestir *et al.*, 1991; Nguyen, 1978), but to extend the underlying semantic background in comparison with pure random set theory (Gebhardt and Kruse, 1993c). Related to this origin of possibility distributions, the major aim of our paper is to develop concepts for an efficient induction of *possibilistic networks* from data, which are families of possibility distributions on low-dimensional subspaces of the universe of discourse, induced by a dependency hypergraph or a directed acyclic graph (DAG) of causal dependencies. Aspects of knowledge propagation (focusing) in possibilistic networks are not addressed here, but have been published elsewhere (Gebhardt and Kruse, 1995b). For a comparison with probabilistic networks, see (Gebhardt and Kruse, 1995a).

Following our introductory remarks, this paper is organized as follows. Section 2 proposes a general framework for the interpretation of databases with imprecise cases. Our investigations start with the special problem of structure identification in relational data. We present an approach which is rather oriented at knowledge-based reasoning in causal constraint networks than the traditional decomposition techniques from database theory. In particular, given a relation R as an imprecise specification of ω_0, the basic idea is to search for a DAG G that minimizes the expected amount of additional information needed beyond R, when the deductive reasoning scheme induced by G is applied to identify or to reject an element of the universe of discourse as the current object state ω_0. We present a theorem for computing such an optimal DAG and an efficient algorithm for approximating it relative to a given class of DAGs. Using the Greedy search method, the algorithm runs in time $O(mn^2r^{k+1})$, where m is the number of cases ($|R| = m$), n the number of variables, r the maximum cardinality of the single domains, and k the maximum number of parents of each node in the DAG. This is a convenient result, since the underlying learning problem is known as being NP-hard, so that the consideration of good heuristics is unavoidable. In Section 3 we deal with an information-compressed possibilistic representation of a database of sample cases and show in which way the results obtained for the relational framework can easily be transferred to the more general possibilistic setting without essentially affecting efficiency conditions. Note that the approach chosen of course does not coincide with viewing a DAG as a directed independence graph in terms of standard graphical modelling (Whittaker, 1990). Our main purpose is to define an adequate heuristic for the efficient automated construction of a DAG G in order to find a good decomposition of a possibility distribution with respect to the attached hypergraph $H(G)$, induced by the moral graph of G. Theoretical investigations on the concept of a possibilistic independence graph, the proof of the possibilistic counterparts of the well-known probabilistic factorization theorems of decomposition, and inducing tightest hypertree decompositions of possibility distributions from data, can be found in (Gebhardt and Kruse, 1996a; 1996b). Section 4 concludes the paper with a summary, a discussion, and a report of an experiment that demonstrates the successful application of our approach to the reconstruction of a non-trivial 22-node network from a given database with 747 cases.

2. Structure Identification in Relational Data

2.1. Basic Issues

Starting with the presentation of a formal framework for the interpretation of a database of imprecise sample cases, the aim of this subsection is to address some basic issues concerning the special problem of structure identification in relational data. As an extension of the introductory example shown in Table 1 to the general case, let $\mathrm{Obj}(X_1, \ldots, X_n)$ be an *object type* of interest, characterized by a set $V = \{X_1, \ldots, X_n\}$ of *variables (attributes)* with finite domains $\Omega^{(i)} = \mathrm{Dom}(X_i)$, $i = 1, \ldots, n$.

The *precise specification* of a current object state of this type is then formalized as a tuple $\omega_0 = (\omega_0^{(1)}, \ldots, \omega_0^{(n)})$, taken from the *universe of discourse* $\Omega = \Omega^{(1)} \times \ldots \times \Omega^{(n)}$. Any subset $R \subseteq \Omega$ can be used as a *set-valued specification* of ω_0, which consists of all states that are possible candidates for ω_0. R is therefore called *correct* for ω_0, if and only if $\omega_0 \in R$. R is called *imprecise*, iff $|R| > 1$, *precise*, iff $|R| = 1$, and *contradictory*, iff $|R| = 0$.

Suppose that general knowledge about the dependencies among the variables is available in the form of a database $\mathcal{D} = (D_j)_{j=1}^{m}$ of sample cases. Each case D_j is interpreted as a (set-valued) correct specification of the previously observed representative object state $\omega_j = (\omega_j^{(1)}, \ldots, \omega_j^{(n)})$.

Supporting imprecision (non-specificity) consists in stating $D_j = D_j^{(1)} \times \ldots \times D_j^{(n)}$, where $D_j^{(i)}$ denotes a non-empty subset of $\Omega^{(i)}$. We assume that $\omega_j^{(i)} \in D_j^{(i)}$ is satisfied, but no further information about any preferences among the elements in $D_j^{(i)}$ is given. When the cases in $\mathcal{D}$ are applied as an imperfect specification of the current object state ω_0, then *uncertainty* concerning ω_0 occurs in the way that the underlying frame conditions, under which a sample state ω_j has been observed and which we call the *context* c_j of ω_j, may only for some of the cases coincide with the context on which the observation of ω_0 is based. A complete description of context c_j depends on the physical frame conditions of ω_j, but is also influenced by the frame conditions of observing ω_j by a human expert, a sensor, or any other observation unit. For the following consideration, we make some assumptions on the relationships between contexts and context-dependent specifications of object states. In particular, we suppose that our knowledge about ω_0 can be represented by an *imperfect specification*

$$\begin{aligned} \Gamma &= (\gamma, P_C) \\ C &= \{c_1, \ldots, c_m\} \\ \gamma : C &\to \mathfrak{P}(\Omega) \\ \gamma(c_j) &= D_j, \qquad j = 1, \ldots, n \end{aligned}$$

with C denoting the set of contexts, $\gamma(c_j)$ the context-dependent set-valued specification of ω_j, P_C a probability measure on C, and $\mathfrak{P}(\Omega)$ the power set of Ω. $P_C(\{c\})$ quantifies the probability of occurrence of context $c \in C$. If all contexts are equally representative and thus equally likely, then P_C should be the uniform distribution on C.

We suppose that C can be described as a subset of logical propositions. The mapping $\gamma : C \to \mathfrak{P}(\Omega)$ indicates the assumption that there is a functional dependency of the sample cases from the underlying contexts, so that each context c_j

determines uniquely its set-valued specification $\gamma(c_j) = D_j$ of ω_j. It is reasonable to state that $\gamma(c_j)$ is correct for ω_j (i.e. $\omega_j \in \gamma(c_j)$) and of *maximum specificity*, which means that no proper subset of $\gamma(c_j)$ is guaranteed to be correct for ω_j with respect to context c_j. Related to the current object state of interest, specified by the (unknown) value $\omega_0 \in R$, and observed in a new context c_0, any c_j in C is adequate for delivering a set-valued specification of ω_0, if c_0 and c_j, formalized as logical propositions, are not contradicting. Intersecting the context-dependent set-valued specifications $\gamma(c_j)$ of all contexts c_j that do not contradict c_0, we obtain the most specific correct set-valued specification of ω_0 with respect to γ.

The idea of using set-valued mappings on probability fields in order to treat uncertain and imprecise data refers to similar random-set-like approaches that were suggested, for instance, in (Dempster, 1968; Kampe de Feriet, 1982; Strassen, 1964). But note that for operating on imperfect specifications in the field of knowledge-based systems, it is important to provide adequate semantics. We addressed this topic in more detail elsewhere (Gebhardt and Kruse, 1993a; 1993b). For the purpose of this paper, the method of carrying out inference based on an imperfect specification Γ of ω_0 is only of theoretical relevance, since simply given a database of (imprecise) samples, it seems to be out of reach and in some sense also not intended to provide complete descriptions of contexts. In Section 3 we will investigate how a *possibilistic* interpretation of a database of sample cases supports well-founded information-compressed reasoning without perfect knowledge of context descriptions.

In this section, we confine to a more elementary situation where the database $\mathcal{D} = (D_j)_{j=1}^m$ consists of pairwise disjoint cases of precise data. Formally speaking, we state the following presuppositions:

(R1) $\quad D_j \neq D_k \quad$ for all $\quad j, k \in \{1, \ldots, m\},\ j \neq k$

(R2) $\quad D_j = \{r_j\} \quad$ for all $\quad j \in \{1, \ldots, m\}$

Additionally, as the first approach to structure identification in relational data, we do not incorporate any uncertainty concept in our considerations. In this restricted framework, there is of course no need for interpreting $\mathcal{D}$ as an imperfect specification $\Gamma = (\gamma, P_C)$ of ω_0, but it is sufficient to recognize that $\mathcal{D}$ induces the relation $R = \{r_1, \ldots, r_m\} \subseteq \Omega$ that reflects the set of all observed dependencies among the values of variables in V.

The detailed description of the underlying contexts as well as the frequency of the occurrence of the same cases in $\mathcal{D}$ is ignored.

All relationships encoded by the tuples contained in $\Omega \backslash R$ are regarded as being impossible for a correct specification of ω_0. R serves as the most specific set-valued specification of ω_0, which means that $\omega_0 \in R$ holds for sure, but $\omega_0 \in R'$ is not guaranteed for any proper subset R' of R. In order to make the dependencies contained in R more evident, we have to establish a meaningful structure of R. This problem is well-known in database theory, namely finding a *decomposition* $\mathcal{E}$ of R into lower-dimensional, more basic dependencies (Maier, 1983; Ullman, 1988; 1989).

From a purely qualitative point of view, a decomposition can be represented with the aid of a *dependency hypergraph* (Berge, 1976) $H = (V, \mathcal{E})$, where V is interpreted as a finite set of *vertices* and $\mathcal{E}$ as a set of *hyperedges*, where each of them is a subset of V. Furthermore, we suppose that the following conditions are satisfied:

$$\text{(H1)} \quad E \neq \emptyset \text{ for all } E \in \mathcal{E}$$

$$\text{(H2)} \quad \bigcup_{E \in \mathcal{E}} E = V$$

$$\text{(H3)} \quad E \not\subset E' \text{ for all } E, E' \in \mathcal{E}$$

Every hyperedge $E \in \mathcal{E}$ reflects a significant dependency among the values of the variables contained in E. Condition (H1) avoids empty hyperedges, (H2) ensures that all variables are contained in the dependency structure, and (H3) assumes that hyperegdes are *reduced (skeletal)*, i.e. no hyperedge E is a proper subset of another hyperedge E', since otherwise E' would already cover the dependencies that can be expressed by E.

From a quantitative point of view, a dependency hypergraph $H = (V, \mathcal{E})$, when applied to a relation R, induces a *constraint network* $\mathcal{N}_H(R)$ over V, which is defined as the family of nonempty relations

$$R_E \stackrel{\text{def}}{=} \Pi_E^V(R)$$

with Π_E^V denoting the pointwise projection from Ω^V onto Ω^E. Let $\Omega^{\{v\}}$ be the domain of the variable $v \in V$. If $W \subseteq V$ is an arbitrary subset of variables, then

$$\Omega^W \stackrel{\text{def}}{=} \begin{cases} \times_{v \in W} \Omega^{\{v\}}, & \text{if} \quad W \neq \emptyset \\ \{\varepsilon\}, & \text{if} \quad W = \emptyset \end{cases}$$

is defined as the product of their domains, where the empty tuple ε is the only element of $\Omega^{\emptyset}$.

Since $\mathcal{N}_H(R)$ specifies local dependencies among the values of the variables in $E \in \mathcal{E}$, $\mathcal{N}_H(R)$ may be less informative than R. In particular, defining

$$\text{rel}\left(\mathcal{N}_H(R)\right) \stackrel{\text{def}}{=} \left\{\omega \in \Omega \mid \forall E \in \mathcal{E} : \Pi_E^V(\omega) \in R_E\right\}$$

as the set of all global dependencies in Ω that can be derived from $\mathcal{N}_H(R)$, we obtain

$$R \subseteq \text{rel}\left(\mathcal{N}_H(R)\right)$$

Structure identification of R is the task of finding a dependency hypergraph $H = (V, \mathcal{E})$ such that

$$R = \text{rel}\left(\mathcal{N}_H(R)\right)$$

holds. The induced constraint network $\mathcal{N}_H(R)$ is then called a *lossless join decomposition* of R which *describes* or *represents* R.

Unfortunately there are some complexity problems concerning structure identification in relational data. If we are given a hypergraph H, then only in the cases where H is tractable (for instance, if H is a hypertree), one can (tractably) decide whether $\text{rel}(\mathcal{N}_H(R)) = R$. On the other hand, lossless join decomposition of a relation into a structure taken from a *class* of dependency hypergraphs turns out to be a harder task, which is presumably intractable even in the cases where each individual member of the class is tractable (Dechter and Pearl, 1992).

From database theory it is known that given a relation R and any arbitrary hypergraph H, deciding whether $\text{rel}(\mathcal{N}_H(R)) = R$, is NP-hard. A conjecture in (Dechter and Pearl, 1992) claims that this result can be extended to the complete hypergraphs

$$H_k = (V, \mathcal{E}_k), \qquad \mathcal{E}_k = \left\{E \subseteq \mathfrak{P}(V) \mid |E| = k\right\}$$

Given a relation R and an integer k, we are therefore in general not in a position to decide whether $\text{rel}(\mathcal{N}_{H_k}(R)) = R$ in polynomial time.

At first glance, these results are not very encouraging, and there are only a few efficient structure identification algorithms for quite special classes of constraint networks. As an example, in the case of tree-structured constraint networks, it is possible to use an algorithm which determines whether a given relation R has a lossless join decomposition, and if the answer is positive, it identifies the topology of the corresponding tree in $O((|R| + \log_2 n)n^2)$ time (Dechter, 1990). Assuming a uniform distribution on R, the underlying approach is based on a procedure for learning the maximum-weight spanning tree of a tree-structured Bayesian network from data (Chow and Liu, 1968). It applies the *Kullback-Leibler cross-entropy* (Kullback and Leibler, 1951; Shore and Johnson, 1980) for probability distributions in order to quantify the strength of dependency between any two variables in the network.

Nevertheless, as a consequence of the above-mentioned crucial complexity problems in the general case, structure identification algorithms for relational data require good heuristic search methods to be efficient. Since there is a very close relationship between constraint satisfaction problems and the recognition of lossless join decompositions in relational databases, which is due to the common underlying structure, namely a hypergraph, some results obtained in one of the two fields can lead to an improvement in the other field. Recently some work has been done to make these relationships explicit, and to show how a constraint satisfaction problem may be decomposed into a number of subproblems that support the development of heuristic solution strategies with an acceptable worst-case complexity bound (Dechter and Pearl, 1989; Gyssens *et al.*, 1994).

2.2. A New Approach to Structure Identification

We now want to investigate the problem of structure identification in relational data from a different point of view, which is more oriented to reasoning in causal networks than dependency hypergraph representations. We refer to the situation when dependencies among the variables can be represented by a *directed acyclic graph (DAG)* $G = (V, E)$ with V as a finite set of *nodes* and $E \subseteq V \times V$ as a set of *arcs*. Any arc $(v, v') \in E$ signifies the presence of a direct causal dependency of the values of the variable v' on the values of the variable v. Since there are no cyclic dependencies in a DAG, it can be seen as a scheme for deductive reasoning from causes to effects. In the following, we shall study the corresponding reasoning process in more detail.

Definition 1. Let $V = \{X_1, \ldots, X_n\}$ be a *set of variables*, $\Omega^{(i)} = \text{Dom}(X_i)$, $i = 1, \ldots, n$, their finite *domains*, and $\Omega = \Omega^{(1)} \times \ldots \times \Omega^{(n)}$ their common *universe of discourse*. Furthermore let $R \subseteq \Omega$ be a nonempty relation and $G = (V, E)$ a DAG. For any $W \subseteq V$ and $X \in V \backslash W$, define

$$\varphi\Big[R; W \to X\Big] \quad : \quad \Omega^W \to \mathfrak{P}(\Omega^{\{X\}})$$

$$\varphi\Big[R; W \to X\Big](\omega) \stackrel{\text{def}}{=} R\Big(X \mid W = \omega\Big), \text{ where}$$

$$R\Big(X \mid W = \omega\Big) \stackrel{\text{def}}{=} \Pi^V_{\{X\}}\Big(\{\omega' \in R \mid \Pi^V_W(\omega') = \omega\}\Big)$$

For any $X \in V$, let

$$\text{par}_G(X) \stackrel{\text{def}}{=} \Big\{Y \in V \mid (Y, X) \in E\Big\}$$

denote the *set of all parent nodes of* X. Then the family

$$\mathcal{N}_G(R) \stackrel{\text{def}}{=} \Big(\varphi\big[R; \text{par}_G(X) \to X\big]\Big)_{X \in V}$$

is called the *causal constraint network* induced by G and R.

If $\text{par}_G(X) \neq \emptyset$, then $R(X \mid \text{par}_G(X) = \omega)$ is the set of those values of effect X that remain possible, if the direct causes $Y \in \text{par}_G(X)$ of X are *instantiated* with the values $\Pi^{\text{par}_G(X)}_{\{Y\}}(\omega)$, which means that all the variables in $\text{par}_G(X)$ are known. Referred to the current object state ω_0, these instantiations induce that the equality $\omega = \Pi^V_{\text{par}_G(X)}(\omega_0)$ holds. $R(X \mid \text{par}_G(X) = \omega)$ is the most specific correct set-valued specification of $\Pi^V_{\{X\}}(\omega_0)$, i.e. $R(X \mid \text{par}_G(X) = \omega)$ is correct for $\Pi^V_{\{X\}}(\omega_0)$, but this correctness property is not ensured for any proper subset of $R(X \mid \text{par}_G(X) = \omega)$.

If $\text{par}_G(X) = \emptyset$, then the variable X does not occur as an effect of any other variable. In this case $\varphi[R; \emptyset \to X]$ collapses to a function with no arguments, and we obtain $R(X \mid \text{par}_G(X) = \varepsilon) = \Pi^V_{\{X\}}(R)$, which in fact is the set of all values of X that can be derived from R without any instantiation.

Example 1. Let $V = \{X_1, X_2\}$ be a set of two variables (attributes) that describe the state of some system. Let $\Omega^{(1)} = \text{Dom}(X_1) = \{0, 1\}$, $\Omega^{(2)} = \text{Dom}(X_2) = \{0, 1, 2\}$ the underlying domains, and $\Omega = \Omega^{(1)} \times \Omega^{(2)} = \{(0,0), (0,1), (0,2), (1,0), (1,1), (1,2)\}$ the common universe of discourse. Additionally, let $R = \{(0,0), (0,1), (1,1), (1,2)\}$ be the most specific correct set-valued specification of the current state ω_0 of interest. Suppose that the dependencies among X_1 and X_2 are represented with the aid of the DAG $G = (V, E)$, $E = \{(X_1, X_2)\}$, saying that X_1 can be viewed as a direct cause of X_2. G and R induce the causal constraint network

$$\mathcal{N}_G(R) = \Big(\varphi[R; \emptyset \to X_1], \varphi[R; \{X_1\} \to X_2]\Big)$$

which indicates how knowledge about X_1 restricts the possible values of X_2.

Let $\omega_0 = (\omega_0^{(1)}, \omega_0^{(2)})$, and assume that $\omega_0^{(1)} = 0$ is known. With respect to $\mathcal{N}_G(R)$ we obtain

$$\varphi\Big[R; \text{par}(X_2) \to X_2\Big](0) = \varphi\Big[R; \{X_1\} \to X_2\Big](0) = \{0,1\}$$

as the most specific correct set-valued specification of $\omega_0^{(2)}$ that follows from R and the additional information $\omega_0^{(1)} = 0$. We thus infer that $\omega_0^{(2)} = 0$ or $\omega_0^{(2)} = 1$, while $\omega_0^{(2)} = 2$ has to be rejected.

Note that the specification $\{0,1,2\}$ is also correct for $\omega_0^{(2)}$ but that it does not satisfy the maximum specificity criterion. On the other hand, any proper subset of $\{0,1\}$ may not be correct for $\omega_0^{(2)}$. Choosing, for example, $\{0\} \subset \{0,1\}$ and $\omega_0 = (0,1)$, the condition $\omega_0 \in R$ and $\omega_0^{(1)} = 0$ holds, but the specification $\{0\}$ is not correct for $\omega_0^{(2)}$. ■

Using the concept of a causal constraint network, we now establish the above-mentioned deductive reasoning scheme. Let $<$ be a topological ordering of V, i.e. all nodes $v, v' \in V$ with $v < v'$ satisfy the condition $(v', v) \notin E$. Suppose that $v_1 < v_2 < \ldots < v_n$ reflects this ordering. For any $\omega \in \Omega$ we want either to identify ω as the current object state ω_0 ($\omega = \omega_0 \in R$) or to verify $\omega \neq \omega_0$ and $\omega \notin R$ by marking those variables v_i, $i = 1, \ldots, n$, for which $\Pi^V_{\{v_i\}}(\omega)$ turns out to contradict the available general knowledge of all possible object states encoded by R, given the instantiations $\omega^{\{Y\}} = \Pi^V_{\{Y\}}(\omega)$ for all direct causes (parent nodes) $Y \in \text{par}_G(v_i)$.

Scheme 1. *Pseudocode of a deductive reasoning scheme for identifying or rejecting any $\omega \in \Omega$ as the current object state ω_0, given a causal constraint network $\mathcal{N}_G(R)$.*

for $i := 1$ **to** n **do begin**

assign $X := v_i$ and calculate $\Phi_i := R\left(X \mid \text{par}_G(X) = \Pi^V_{\text{par}_G(X)}(\omega)\right)$;
(the topological ordering of V ensures that $\text{par}_G(X)$ is a set of already instantiated variables)

if $\Phi_i \neq \emptyset$

then provide further information such that $\Pi^V_{\{X\}}(\omega_0)$ is identified within the set Φ_i of remaining possible instantiations of X, or $\Pi^V_{\{X\}}(\omega) \neq \Pi^V_{\{X\}}(\omega_0)$ is recognized

end

if $(\Phi_i = \emptyset)$ **or** $(\Pi^V_{\{X\}}(\omega) \neq \Pi^V_{\{X\}}(\omega_0))$

then mark X as a variable where an erroneous instantiation has been detected $(\omega \neq \omega_0)$

end

end

The amount of information that has to be added in order to identify $\Pi^V_{\{v_i\}}(\omega_0)$ within the set Φ_i of possible alternatives can be quantified by $H(\Phi_i)$, where H denotes the *Hartley measure of information* (Hartley, 1928).

For any non-empty finite set A,

$$H(A) \stackrel{\text{def}}{=} \log_2 |A|$$

is the number of elementary propositions (measured in bits), whose truth values must be determined for the specification and thus the identification of a single element in the reference set A.

Note that in the case $\Phi_i = \emptyset$, there is no need for adding further information, since no possible alternative instantiations of v_i exist. We therefore define

$$H(\emptyset) \stackrel{\text{def}}{=} 0$$

Example 1. (continued) Let $\omega = (\omega_0^{(1)}, \omega_0^{(2)}) = (0, 1)$. We want to identify ω as the current object state $\omega_0 = (\omega^{(1)}, \omega^{(2)})$. The application of Scheme 1 to the node ordering $X_1 < X_2$ yields:

1. $\Phi_1 = R(X_1 \mid \text{par}_G(X_1) = \varepsilon) = \Pi^V_{\{X_1\}}(R) = \Omega^{(1)} = \{0, 1\}$, where $\text{par}_G(X_1) = \emptyset$ and $\Pi^V_{\emptyset}(\omega) = \varepsilon$.

2. The amount of information needed in order to identify $\omega_0^{(1)} = \omega^{(1)} = 0$ in the set Φ_1 of possible alternatives is $H(\Phi_1) = \log_2 |\Phi_1| = 1$ bit.

3. $\Phi_2 = R(X_2 \mid \text{par}_G(X_2) = 0) = \{0, 1\} \subset \Omega^{(2)} = \{0, 1, 2\}$, where $\text{par}_G(X_1) = \{X_1\}$ and $\Pi^V_{\{X_1\}}(\omega) = 0$.

4. The additional amount of information for identifying $\omega_0^{(2)} = \omega^{(2)} = 1$ in Φ_2 is $H(\Phi_2) = \log_2 |\Phi_2| = 1$ bit.

Hence, using the deductive reasoning scheme induced by G, the total amount of information needed in order to identify ω as the current object state ω_0 is 2 bits.

Let $\omega = (\omega^{(1)}, \omega^{(2)}) = (0, 2)$. Then $\Pi^V_{\{X_2\}}(\omega) = 2 \notin \Phi_2 = \{0, 1\}$, and we mark X_2 as a variable where an erroneous instantiation has been detected. This implies $\omega \notin R$ and therefore $\omega \neq \omega_0$, so that ω is rejected as the current object state. ∎

Based on Hartley information, we measure the nonspecificity of $\mathcal{N}_G(R)$ with respect to $\omega \in \Omega$, namely the total amount of additional information (beyond R) that is necessary either to identify ω as the current object state ($\omega = \omega_0$) by carrying out inference in the presented deductive reasoning scheme, or to mark all variables for which the reasoning process provides a contradiction ($\omega \neq \omega_0$, $\omega \notin R$). We summarize these ideas in the following definition.

Definition 2. Let $G = (V, E)$ be a DAG, $R \subseteq \Omega$ a non-empty relation, and $\mathcal{N}_G(R) = (\varphi[R; \mathrm{par}_G(X) \to X])_{X \in V}$ their causal constraint network. Then, for any $\omega \in \Omega$,

$$\mathrm{Nonspec}[\mathcal{N}_G(R)](\omega) \stackrel{\mathrm{def}}{=} \sum_{X \in V} H\Big(R(X \mid \mathrm{par}_G(X) = \Pi^V_{\mathrm{par}_G(X)}(\omega))\Big)$$

is called the *nonspecificity of* $\mathcal{N}_G(R)$ *w.r.t.* ω.

While interpreting $\mathrm{Nonspec}[\mathcal{N}_G(R)] : \Omega \to \mathbb{R}$ as a uniformly distributed random variable,

$$E\Big(\mathrm{Nonspec}[\mathcal{N}_G(R)]\Big) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \mathrm{Nonspec}[\mathcal{N}_G(R)](\omega)$$

is the *expected nonspecificity of* $\mathcal{N}_G(R)$.

The concept of the expected nonspecificity of a causal constraint network is helpful for inducing an optimal DAG G that minimizes $E(\mathrm{Nonspec}[\mathcal{N}_G(R)])$ relative to R and a chosen class of DAGs. Our assumption that $\mathrm{Nonspec}[\mathcal{N}_G(R)]$ is a uniformly distributed random variable coincides with the *a priori* state of knowledge that $\omega_0 = \omega$ is equally likely for all $\omega \in \Omega$.

The following algorithm constructs a DAG G that is optimal relative to a class of DAGs that we will specify in more detail afterwards.

Algorithm 1.

Input: A non-empty relation $R \subseteq \Omega$ and a number $k \in \mathbb{N}$.

Output: A DAG $G = (V, E)$ for the induction of a causal constraint network $\mathcal{N}_G(R)$ that satisfies $|\mathrm{par}_G(X)| \leq k$ for all $X \in V$ and reflects significant causal dependencies in R.

$V^* := V$; $E := \emptyset$;

while $V^* \neq \emptyset$ **do begin**

select a node $X \in V^*$ and a set $W \subseteq V$ of possible parent nodes ($X \notin W$, $|W| \leq k$) such that the DAG property of G is not affected, if the arcs (Y, X) for all $Y \in W$ are added to G, and the quantity

$$m(W, X) = \frac{1}{|\Omega^W|} \sum_{\omega \in \Omega^W} H\Big(R(X \mid W = \omega)\Big)$$

is minimized.

$E := E \cup \{(Y, X) \mid Y \in W\}$;

$V^* := V^* \backslash \{X\}$

end

The numbers $m(W, X)$ quantify the *degree of independence* of X from the variables in W. In this context we want to distinguish two different kinds of independence concerning the treatment of imprecise data, namely the concepts of *strong* and *weak independence.*

X is called *strongly independent* of W relative to R, iff R does not help us to leave the state of total ignorance about the value of X (i.e. $\Pi^V_{\{X\}}(\omega_0) \in \Omega^{\{X\}}$) when any additional knowledge about the values of the variables in W becomes available. On the other hand, X is called *weakly independent* of W relative to R, iff learning something about the values of the variables in W does not further restrict the set of possible values of X.

Formally speaking, X is strongly independent of W relative to R, iff

$$\forall \omega \in \Omega^W : R\Big(X \mid W = \omega\Big) = \Omega^{\{X\}}$$

and X is weakly independent of W relative to R, iff

$$\forall \omega \in \Omega^W : R\Big(X \mid W = \omega\Big) = \Pi^V_{\{X\}}(R)$$

Strong independence thus implies weak independence, whereas the converse is not true.

If we define

$$m^*(W, X) \stackrel{\text{def}}{=} \begin{cases} \dfrac{m(W, X)}{H(\Omega^{\{X\}})}, & \text{iff} \quad |\Omega^{\{X\}}| > 1 \\ 1, & \text{iff} \quad |\Omega^{\{X\}}| = 1 \end{cases}$$

then the quantity $m^*(W, X)$ is the degree of strong independence of X from the variables in W relative to R.

The minimum value $m^*(W, X) = 0$ reflects a functional dependence, whereas the maximum value $m^*(W, X) = 1$ characterizes strong independence as the weakest form of relational dependence.

In a similar way,

$$m_*(W, X) \stackrel{\text{def}}{=} \begin{cases} \dfrac{m(W, X)}{H(\Pi^V_{\{X\}}(R))}, & \text{iff} \quad |\Pi^V_{\{X\}}(R)| > 1 \\ 1, & \text{iff} \quad |\Pi^V_{\{X\}}(R)| = 1 \end{cases}$$

quantifies the degree of weak independence of X from the variables in W relative to R.

Note that weak independence coincides with strong independence when referred to the restricted universe of discourse

$$\Omega' \stackrel{\text{def}}{=} \times_{i=1}^n \Pi^V_{\{X_i\}}(R)$$

From this point of view, there is in fact only one underlying concept of independence, which may be applied to different universes of discourse. Whereas weak independence refers to the dynamics of knowledge expansion, strong independence is rather oriented to the static original universe of discourse chosen for the initial state of total ignorance about ω_0. For this reason, $m_*(W, X)$ should be preferred in the case where we are searching for a decomposition of a relation, but using $m^*(W, X)$ has obvious advantages when we want to quantify the nonspecificity of a relation relative to Ω.

For our purposes, we do not need any of the two independence concepts in an explicit way, but we only use the weights $m(W, X)$. The following theorem shows that these quantities are adequate in order to determine an optimal DAG for a given relation R.

Theorem 1. *Using the weights $m(W, X)$ as defined in Algorithm 1, we have*

$$E\Big(\text{Nonspec}[\mathcal{N}_G(R)]\Big) = \sum_{X \in V} m\Big(\text{par}_G(X), X\Big)$$

With respect to the above theorem, it is quite easy to check the optimality of a DAG relative to R. On the other hand, it does not provide an efficient procedure to construct an optimal DAG. Therefore we study the heuristic search method of Algorithm 1 concerning the optimality of the output DAGs. As a preparation we introduce an additional concept:

For any DAG $G = (V, E)$, let $\text{cons}(G)$ denote the transitive hull of E, namely the smallest subset of $V \times V$ such that the conditions

(T1) $E \subseteq \text{cons}(G)$

(T2) $(X, Y) \in \text{cons}(G)$ and $(Y, Z) \in \text{cons}(G)$ implies $(X, Z) \in \text{cons}(G)$

hold.

The set $\text{cons}(G)$ specifies a partial ordering of V and thus all constraints for a linear node ordering that follow from the topology of G. We say that a DAG $G' = (V, E')$ *satisfies* the node ordering constraints of G, iff $\text{cons}(G) \subseteq \text{cons}(G')$.

Let

$$\mathcal{G}_k(V) \stackrel{\text{def}}{=} \Big\{G^* = (V, E^*) \mid G^* \text{ DAG and } \forall X \in V : |\text{par}_{G^*}(X)| \leq k\Big\}$$

The next Theorem clarifies that the optimality of any output DAG $G \in \mathcal{G}_k(V)$ of Algorithm 1 is relative to the subclass of $\mathcal{G}_k(V)$ that satisfies the node ordering constraints of G.

Theorem 2. *Let G be a DAG that is constructed by application of Algorithm 1. Then*

$$E\Big(\text{Nonspec}[\mathcal{N}_G(R)]\Big) = \min\Big\{E\big(\text{Nonspec}[\mathcal{N}_{G'}(R)]\big) \mid G' \in \mathcal{G}_k(V) \text{ and } \text{cons}(G) \subseteq \text{cons}(G')\Big\}$$

Note that the time complexity of an efficient deterministic implementation of Algorithm 1 is $O(mn^{k+1}r^{k+1})$ and therefore it is polynomial with respect to n, where r denotes the maximum cardinality of the single domains. The fact that optimality is only reached relative to a specific class of DAGs is not surprising, since we already emphasized that, due to the complexity problems of a general solution, good heuristics (as proposed by our strategy of minimizing the expected nonspecificity of deductive reasoning in a dependency DAG) have to be considered in order to reach efficiency.

Any DAG $G = (V, E)$ that is obtained as an output of Algorithm 1 or with the aid of Theorem 1 can be used to define a dependency hypergraph $H(G)$, induced by the moral graph of G, which is

$$H(G) = (V, \mathcal{E}_G), \qquad \mathcal{E}_G \stackrel{\text{def}}{=} \Big\{\{X\} \cup \text{par}_G(X) \mid X \in V\Big\}$$

and thus a constraint network $\mathcal{N}_{H(G)}(R)$ such that $R \subseteq \text{rel}(\mathcal{N}_{H(G)}(R))$ holds. Hence, if R is a correct (set-valued) specification of ω_0, then $\text{rel}(\mathcal{N}_{H(G)}(R))$ is correct for ω_0, too. For this reason $\text{rel}(\mathcal{N}_{H(G)}(R))$ may serve as an approximation of R, and the distance $|\text{rel}(\mathcal{N}_{H(G)}(R))| - |R|$ as a measure of the approximation quality. It is near at hand that there should be a strong relationship between this distance and the expected amount $E(\text{Nonspec}[\mathcal{N}_G(R)])$ of information to be added to $\mathcal{N}_G(R)$ in order to identify or to reject any element $\omega \in R$ as the current object state. But it is not clear to us at the moment in which cases a DAG G minimizes $E(\text{Nonspec}[\mathcal{N}_G(R)])$ when $\mathcal{N}_{H(G)}(R)$ is a lossless join decomposition of R, because the concept of expected nonspecificity is more oriented to reasoning tasks than structure identification. It is also not obvious in which way decreasing values $E(\text{Nonspec}[\mathcal{N}_G(R)])$ lead to relations $\text{rel}(\mathcal{N}_{H(G)}(R))$ that are tighter approximations of R. Presumably, optimal results can only be reached when $H(G)$ is a hypertree.

Nevertheless, the results obtained in the following example, and especially the success of using an extension of the presented concepts to the more complex problem of inducing structure in the possibilistic setting that we will introduce in the next section, confirm the presumable power of our approach.

Also consider that Algorithm 1 is based on a simple *ad-hoc* search method that may be improved in various directions, one of which refers to the node selection process, when the set of nodes X that minimizes $m(W, X)$ consists of more than only one node. Anyway, other techniques of minimizing $E(\text{Nonspec}[\mathcal{N}_G(R)])$ can be used.

Constructing tight approximations of R from an efficiency point of view leads to the development of Greedy search algorithms. One example is to compute $W_l(W)$ for all $X \in V$, satisfying

$$m\Big(W_l(X), X\Big) = \min\Big\{m(W, X) \mid W \subseteq V, |W| \leq l, W_{l-1}(X) \subseteq W\Big\}$$

where $W_0(X) \stackrel{\text{def}}{=} \emptyset$, starting with $l = 1$, and terminating the iteration after calculating $W_k(X)$ without affecting the DAG property of

$$G_l = (V, E_l), \qquad E_l \stackrel{\text{def}}{=} \left\{(Y, X) \mid X \in V \text{ and } Y \in W_l(X)\right\}$$

Note that this algorithm has a nice time complexity of $O(mn^2r^{k+1})$.

An additional strategy of a Greedy search algorithm could avoid an increasing number of parent nodes, whenever the decrease of expected nonspecificity, connected with the addition of one more node, does not exceed a predefined threshold.

We now summarize the main ideas discussed in this section with the aid of a small example of inducing structure from relational data. The calculations carried out in this example are easy to follow. An extension will be presented at the end of Section 3.

Example 2. With respect to our introductory example, let $V = \{X_1, X_2, X_3, X_4, X_5\}$ be a set of binary variables with domains $\Omega^{(i)} = \text{Dom}(X_i) = \{0, 1\}$, $i = 1, \ldots, 5$. Suppose that there is general knowledge about dependencies among the values of these variables encoded as the relation $R \subseteq \Omega \subseteq \{0, 1\}^5$ given in Table 2.

Tab. 2. A dependency relation R.

tuple	X_1	X_2	X_3	X_4	X_5
1	1	0	1	1	1
2	0	1	0	1	1
3	0	1	1	1	0
4	0	1	0	0	1
5	1	0	1	1	0
6	0	1	1	0	1
7	0	1	1	1	1
8	0	1	0	1	0

Let $H = (V, \mathcal{E})$, defined by $\mathcal{E} = \{\{X_1, X_2\}, \{X_2, X_3\}, \{X_2, X_4\}, \{X_4, X_5\}\}$, denote the dependency hypergraph that coincides with the tree-structured undirected graph in Fig. 1. Then H and R constitute the constraint network $\mathcal{N}_H(R) = (R_E)_{E \in \mathcal{E}}$ that consists of the relations shown in Table 3.

$\mathcal{N}_H(R)$ is a lossless join decomposition of R, so that the equality $\text{rel}(\mathcal{N}_H(R)) = R$ holds.

We now want to apply Algorithm 1 in order to get a DAG $G = (V, E)$ as a deductive reasoning scheme for R. We state that G is tree-structured and choose R and $k = 1$ as the inputs of the algorithm. The first step in constructing G is the computation of the quantities $m(W, X)$ for all $W \subseteq V$ and $X \in V$ that satisfy $|W| \leq 1$ and $X \notin W$. The corresponding results are presented in Table 4.

We apply Theorem 1 and realize that $E(\text{Nonspec}[\mathcal{N}_G(R)]) = 2.5$ is satisfied for each DAG G that minimizes the expected nonspecificity of its induced causal

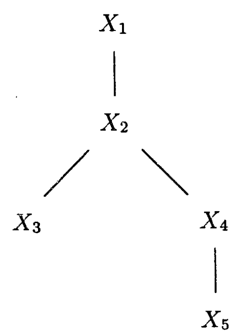

Fig. 1. A tree-structured dependency hypergraph.

Tab. 3. The relations of the constraint network $\mathcal{N}_H(R)$.

$R_{\{X_1,X_2\}}$

X_1	X_2
0	1
1	0

$R_{\{X_2,X_3\}}$

X_2	X_3
0	1
1	0
1	1

$R_{\{X_2,X_4\}}$

X_2	X_4
0	1
1	0
1	1

$R_{\{X_4,X_5\}}$

X_4	X_5
0	1
1	0
1	1

Tab. 4. The quantities $m(W,X)$.

W \ X	X_1	X_2	X_3	X_4	X_5
$\emptyset$	1	1	1	1	1
$\{X_1\}$	•	0	0.5	0.5	1
$\{X_2\}$	0	•	0.5	0.5	1
$\{X_3\}$	0.5	0.5	•	1	1
$\{X_4\}$	0.5	0.5	1	•	0.5
$\{X_5\}$	1	1	1	0.5	•

constraint network $\mathcal{N}_G(R)$ relative to the class $\mathcal{G}_1(V)$ of all tree-structured DAGs. Figure 2 establishes all dependency trees that assume this minimum value.

These trees coincide with the possible outputs of non-deterministic Algorithm 1. One example of a sequence of iteration steps is presented in Table 5, while the resulting DAG G and its attached dependency hypergraph $H(G)$ in Fig. 3.

Note that $H(G)$ is the dependency hypergraph of all possible DAGs that may be delivered by Algorithm 1. $H(G)$ induces a constraint network $\mathcal{N}_{H(G)}(R)$ that equals the lossless join decomposition of Table 3. Hence, the structure identification problem of R has been solved successfully. ■

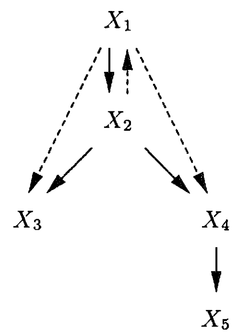

Fig. 2. The topology of all dependency trees G that minimize $E(\text{Nonspec}[\mathcal{N}_G(R)])$.

Tab. 5. Sequence of iteration steps in Algorithm 1.

step	V^*	select X	select W	$m(W,X)$	add to E
1	$\{X_1, X_2, X_3, X_4, X_5\}$	X_1	$\{X_2\}$	0	(X_2, X_1)
2	$\{X_2, X_3, X_4, X_5\}$	X_5	$\{X_4\}$	0.5	(X_4, X_5)
3	$\{X_2, X_3, X_4\}$	X_3	$\{X_2\}$	0.5	(X_2, X_3)
4	$\{X_2, X_4\}$	X_4	$\{X_2\}$	0.5	(X_2, X_4)
5	$\{X_2\}$	X_2	$\emptyset$	1	—

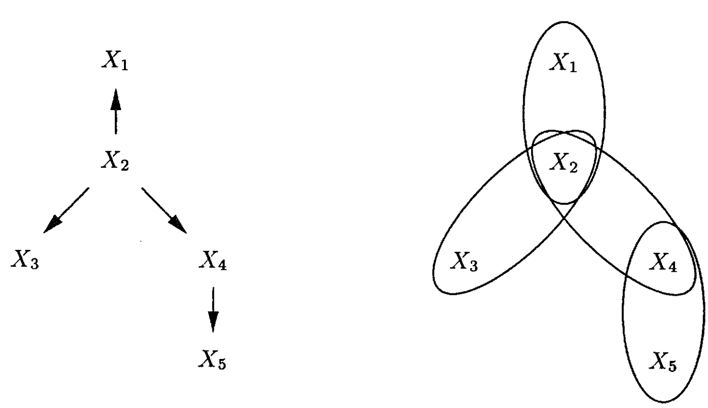

Fig. 3. A tree-structured DAG as an output of Algorithm 1 and its attached dependency hypergraph.

3. Inducing Possibilistic Networks from Data

Our investigations have been so far confined to the problem of structure identification in relational data. This section will extend our approach from relations to databases and from tuples of precise values to imprecise cases.

We already proposed a framework for modelling imprecise and uncertain data based on the concept of an *imperfect specification* $\Gamma = (\gamma, P_C)$ of the current object state $\omega_0 \in \Omega$ of interest, where $\gamma : C \rightarrow \mathfrak{P}(\Omega)$ is a set-valued mapping from a set of consideration contexts to the set of all possible set-valued specifications of object states in Ω, and P_C is a probability measure on C.

In our setting we choose $C = \{c_1, \ldots, c_m\}$ with c_j, $j = 1, \ldots, m$, denoting the context that describes the physical and the observation-related frame conditions that determine $\gamma(c_j)$ as the most specific correct set-valued specification of the sample object state ω_j. $P_C(\{c_j\})$ is the probability of occurrence of c_j and thus it quantifies the weight with which c_j may also serve as the underlying context for specifying ω_0.

When we are only given a database $\mathcal{D} = (D_j)_{j=1}^m$ of sample cases, where $D_j \subseteq \Omega$ is assumed to be a context-dependent most specific specification of ω_j, we are normally not in a position to fully describe the contexts c_j in the form of logical propositions. For this reason, it is convenient to carry out information compression by paying attention to the context-dependent specifications rather than to the contexts themselves. We do not directly refer to $\Gamma = (\gamma, P_C)$, but to its degree of *α-correctness* w.r.t. any ω, which is defined as the total mass of all contexts c_j that yield a correct context-dependent specification $\gamma(c_j)$ of ω_0. If we are given any $\omega \in \Omega$, then $\Gamma = (\gamma, P_C)$ is called *α-correct* w.r.t. ω, iff

$$P_C\Big(\Big\{c \in C \mid \omega \in \gamma(c)\Big\}\Big) \geq \alpha, \qquad 0 \leq \alpha \leq 1$$

Note that, although the use of a probability measure P_C suggests disjoint contexts, we do not make any assumptions about the interrelation of contexts. They may be—with respect to the frame conditions that cause the observations—identical, partially corresponding, or disjoint—we just do not know. We add their weights, because disjoint contexts are the "worst case" in which we cannot restrict the total weight to a smaller value without losing correctness. In this manner, a possibility degree is the upper bound for the total weight of the combined contexts.

Suppose that our only information about ω_0 is the α-correctness of Γ w.r.t. ω_0, without having any knowledge of the description of the contexts in C. Under these restrictions, we are searching for the most specific set-valued specification $A_\alpha \subseteq \Omega$ of ω_0, namely the largest subset of Ω such that α-correctness of Γ w.r.t. ω is satisfied for all $\omega \in A_\alpha$. It easily turns out that the family $(A_\alpha)_{\alpha \in [0,1]}$ consists of all *α-cuts* $[\pi_\Gamma]_\alpha$ of the induced *possibility distribution*

$$\begin{aligned} \pi_\Gamma : \Omega &\rightarrow [0,1] \\ \pi_\Gamma(\omega) &= P_C\Big(\Big\{c \in C \mid \omega \in \gamma(c)\Big\}\Big) \end{aligned}$$

where for any π, taken from the set $\mathrm{POSS}(\Omega)$ of all possibility distributions that can be induced from imperfect specifications w.r.t. Ω, the α-cut $[\pi]_\alpha$ is defined as

$$[\pi]_\alpha = \Big\{\omega \in \Omega \mid \pi(\omega) \geq \alpha\Big\}, \qquad 0 < \alpha \leq 1$$

$$[\pi]_0 = \Omega$$

Note that $\pi_\Gamma(\omega)$ can in fact be viewed as a degree of possibility for the truth of "$\omega = \omega_0$": If $\pi_\Gamma(\omega) = 1$, then $\omega \in \gamma(c)$ holds for all contexts $c \in C$, which means that $\omega = \omega_j$ is possible for all sample object states ω_j, $j = 1, \ldots, m$, so that $\omega_0 = \omega$ should be possible without any restriction.

If $\pi_\Gamma(\omega) = 0$, then $\omega_j = \omega$ has been rejected for ω_j, $j = 1, \ldots, n$, since $\omega \notin \gamma(c_j)$ is true for the set-valued specifications $\gamma(c_j)$ of ω_j. This entails the impossibility of $\omega_0 = \omega$, if the description of the context c_0 for the specification of ω_0 is assumed to be a conjunction of the descriptions of any contexts in C.

If $0 < \pi_\Gamma(\omega) < 1$, then there are contexts that support $\omega_0 = \omega$ as well as contexts that contradict $\omega_0 = \omega$. The quantity $\pi_\Gamma(\omega)$ reflects the maximum possible total mass of contexts that support $\omega_0 = \omega$.

It has to be emphasized that in the special case of precise data (i.e. $|\gamma(c)| = 1$ for all $c \in C$), the possibility distribution π_Γ formally coincides with a probability distribution on Ω, but note that its interpretation is quite different. This becomes obvious when we compare the *probability* $P(A)$ of an event "$\omega_0 \in A$," $A \subseteq \Omega$, with the *possibility* $\Pi(A)$ of the same event: Since $\pi_\Gamma(\omega)$ is the degree of α-correctness of Γ w.r.t. ω, $\Pi(A)$ is defined as the maximum degree of α-correctness of Γ w.r.t. of any element contained in A. Hence, we obtain

$$\Pi(A) \stackrel{\text{def}}{=} \max_{\omega \in A} \pi_\Gamma(\omega)$$

whereas

$$P(A) = \sum_{c \in C : A \cap \gamma(c) \neq \emptyset} P_C\Big(\{c\}\Big)$$

quantifies the total mass of all contexts that support the considered event "$\omega_0 \in A$."

In case $|\gamma(c)| = 1$ holds for all $c \in C$, $P(A)$ can directly be related to the possibility distribution π_Γ:

$$P(A) = \sum_{\omega \in A} \pi_\Gamma(\omega)$$

Note that in recent years several proposals for the semantics of the theory of possibility as a framework for reasoning with uncertain and imprecise data have been made. Among the numerical approaches, we like to mention the epistemic interpretation of fuzzy sets (Zadeh, 1978), the axiomatic view of possibility theory using possibility measures (Dubois and Prade, 1988), one-point coverages of random sets (Hestir *et al.*, 1991; Nguyen, 1978), contour functions of consonant belief functions (Shafer, 1976), falling shadows in set-valued statistics (Wang, 1983), Spohn's theory of epistemic states (Spohn, 1990), and possibility theory based on likelihoods (Dubois *et al.*, 1993). While ignoring the underlying interpretation of contexts, π_Γ formally coincides with one-point coverage of Γ, when it is interpreted as a (not necessarily nested) random set. From a semantics point of view, operating on possibility distributions in our setting may be strongly oriented to the concept of α-correctness. For an extensive presentation of this background of possibility theory, we refer to (Gebhardt and Kruse, 1993b; 1993c). Special aspects of possibility measures for decision making in this framework have been considered in (Gebhardt and Kruse, 1994). We do not go into further details here, but focus our interest on the problem of generalizing the concepts and results presented in the previous section to the possibilistic setting.

As the starting point for the definition of a *possibilistic (causal) network* and its induction from a database $\mathcal{D} = (D_j)_{j=1}^m$ of (imprecise) sample cases, we suppose, just as in Section 2, that all relevant dependencies among the variables in V can be represented in the form of a hypergraph $H = (V, \mathcal{E})$, so that any most specific set-valued specification R of the current object state ω_0 is assumed to have a lossless join decomposition $(R_E)_{E\in\mathcal{E}}$. This property has an important influence on the interpretation of $\mathcal{D}$, caused by the decomposability of $\mathcal{D}$ into a family $(\mathcal{D}_E)_{E\in\mathcal{E}}$ of databases, where each D_E provides sample cases of observed dependencies among the variables contained in the hyperedge E. The database $\mathcal{D}_E = (D_j^E)_{j=1}^m$ consists of the set-valued specifications $D_j^E = \Pi_E^V(D_j)$ of the dependencies $\Pi_E^V(\omega_j)$ that are part of the sample object states ω_j. Given the dependency hypergraph H and the database $\mathcal{D}$, the pair $(\mathcal{D}, H)$ may be applied in order to specify ω_0 imperfectly. Our knowledge about ω_0 can be represented with the aid of a family $(\Gamma_E)_{E\in\mathcal{E}}$ of imperfect specifications $\Gamma_E = (\gamma_E, P_E)$, where $\gamma_E : C_E \to \mathfrak{P}(\Omega^E)$ is defined on the set $C_E = \{c_1^E, \ldots, c_m^E\}$ of those contexts c_j^E that reflect the frame conditions for specifying the sample dependency $\Pi_E^V(\omega_j)$, i.e., $\gamma_E(c_j^E) \overset{\text{def}}{=} D_j^E$.

Under the assumption that the contexts c_j^E in C^E are equally likely, we choose $P_E(c_j^E) \overset{\text{def}}{=} 1/m$ for all $j = 1, \ldots, m$. The family $\mathcal{N}_H(\mathcal{D}) = (\pi_{\Gamma_E})_{E\in\mathcal{E}}$ of the induced possibility distributions π_{Γ_E} is called a *possibilistic (constraint) network* over V.

Stating α_E-correctness of Γ_E w.r.t. $\Pi_E^V(\omega_0)$, we obtain $R_E = [\pi_{\Gamma_E}]_{\alpha_E}$ as the most specific correct set-valued specification of $\Pi_E^V(\omega_0)$ that follows from the interpretation of $\mathcal{D}$ and H. The specifications R_E can be combined with the constraint network $\mathcal{N} = (R_E)_{E\in\mathcal{E}}$ as a lossless join decomposition of rel($\mathcal{N}$), which is the resulting most specific correct set-valued specification of ω_0.

Calculating rel($\mathcal{N}$) requires not only the availability of the database $\mathcal{D}$, but also knowledge of the dependency structure H. In order to induce H from $\mathcal{D}$, we refer to our preparations in Section 2 and assume that the existing dependencies among the variables can be represented by a DAG as a scheme for deductive reasoning from causes to effects. The following definition is the straightforward extension of the concept of a causal constraint network to the concept of a *possibilistic (causal) network.*

Definition 3. Let $V = \{X_1, \ldots, X_n\}$ be a set of variables, $\Omega^{(i)} = \text{Dom}(X_i)$, $i = 1, \ldots, n$, their attached finite domains, and $\Omega = \Omega^{(1)} \times \ldots \times \Omega^{(n)}$ their common universe of discourse. Furthermore, let $\mathcal{D} = (D_j)_{j=1}^m$ be a database of (imprecise) sample cases $D_j = D_j^{(1)} \times \ldots \times D_j^{(n)}$ with $\emptyset \neq D_j^{(i)} \subseteq \Omega^{(i)}$ for $j = 1, \ldots, m$.

Let $\Gamma = (\gamma, P_C)$, determined by $\gamma : C \to \mathfrak{P}(\Omega)$, $C = \{c_1, \ldots, c_m\}$, and $\gamma(c_j) = D_j$, $j = 1, \ldots, m$, be an imperfect specification of the current object state $\omega_0 \in \Omega$ of interest.

Let $G = (V, E)$ be a directed acyclic graph. For any $W \subseteq V$ and $X \in V \backslash W$, define

$$\varphi\Big[\mathcal{D}; W \to V\Big] : \ \Omega \to \mathrm{POSS}\left(\Omega^{\{X\}}\right)$$

$$\varphi\Big[\mathcal{D}; W \to V\Big](\omega) \stackrel{\mathrm{def}}{=} \pi_{\mathcal{D}}\Big(X \mid W = \omega\Big), \text{ where}$$

$$\pi_{\mathcal{D}}\Big(X \mid W = \omega)(\omega'\Big) \stackrel{\mathrm{def}}{=} P_C\Big(\Big\{c \in C \mid \omega \in \Pi_W^V(\gamma(c)) \text{ and } \omega' \in \Pi_{\{X\}}^V(\gamma(c))\Big\}\Big)$$

Then the family $\mathcal{N}_G(\mathcal{D}) = (\varphi[\mathcal{D}; \mathrm{par}_G(X) \to X])_{X \in V}$ is called a *possibilistic causal network* for ω_0, induced by $\mathcal{D}$ and G.

Note that $[\pi_{\mathcal{D}}(X \mid W = \omega)]_\alpha$ is the most specific correct set-valued specification of $\Pi_{\{X\}}^V(\omega_0)$ that follows from $\mathcal{D}$, given the instantiation of the variables in W (i.e. $\Pi_W^V(\omega_0) = \omega$) and the α-correctness of $\Gamma_{W \cup \{X\}}$ w.r.t. $\Pi_{W \cup \{X\}}^V(\omega_0)$.

When choosing an ordering $v_1 < v_2 < \ldots < v_n$ of the nodes in V, the concept of a possibilistic causal network is helpful for specifying an abstract deductive reasoning scheme for the identification or rejection of any $\omega \in \Omega$ as the current object state of interest. It extends Scheme 1 by the additional step of determining for all $Z = \mathrm{par}_G(X) \cup \{X\}$, $X \in V$, the total mass α_Z of all contexts in C_Z whose description does not contradict the description of the context c_0^Z for the specification of the dependency $\Pi_Z^V(\omega_0)$ within the current object state ω_0. Let $\omega' = \Pi_{\mathrm{par}_G(X)}^V(\omega)$ and $\omega'' = \Pi_{\{X\}}^V(\omega)$. Since we assume that the descriptions of the addressed contexts are not available, we only know that $0 \leq \alpha_Z \leq \alpha(\omega', \omega'', X)$ holds, where

$$\alpha(\omega', \omega'', X) \stackrel{\mathrm{def}}{=} \pi_{\mathcal{D}}\Big(X \mid \mathrm{par}_G(X) = \omega'\Big)(\omega'')$$

is the maximum possible correctness degree of Γ_Z w.r.t. $\Pi_Z^V(\omega)$. Applied to any $\omega \in \Omega$, the reasoning scheme works as follows:

Scheme 2. *Pseudocode of a deductive reasoning scheme for identifying or rejecting any $\omega \in \Omega$ as the current object state ω_0, given a possibilistic causal network $\mathcal{N}_G(\mathcal{D})$.*

for $i := 1$ **to** n **do begin**

assign $X := v_i$ and calculate $\pi_i := \pi_{\mathcal{D}}(X \mid \mathrm{par}_G(X) = \Pi_{\mathrm{par}_G(X)}^V(\omega))$;

if $\pi_i \not\equiv 0$

then determine the total mass α_Z of all contexts in C_Z, $Z = \mathrm{par}_G(X) \cup \{X\}$, whose description does not contradict the description of the context c_0^Z for the specification of $\Pi_Z^V(\omega_0)$; (note that Γ_Z is α_Z-correct w.r.t. $\Pi_Z^V(\omega_0)$)

provide further information such that either $\Pi^V_{\{X\}}(\omega_0)$ is identified within the set $\Phi_i := [\pi_i]_{\alpha_Z}$ of remaining possible instantiations of X, or $\Pi^V_{\{X\}}(\omega) \neq \Pi^V_{\{X\}}(\omega_0)$ is recognized.

end;

if $(\pi_i \equiv 0)$ **or** $(\Pi^V_{\{X\}}(\omega) \neq \Pi^V_{\{X\}}(\omega_0))$

then mark X as a variable where an erroneous instantiation has been detected $(\omega \neq \omega_0)$

end

end

Assuming that all choices of α_Z-correctness degrees in the intervals $[0, \alpha(\omega', \omega'', X)]$ are equally likely, we obtain the generalizations of Definition 2, Algorithm 1, Theorem 1, and Theorem 2 to the possibilistic setting.

Definition 4. Let $G = (V, E)$ be a DAG, $\mathcal{D}$ a database of sample cases, and $\mathcal{N}_G(\mathcal{D}) = (\varphi[\mathcal{D}; \mathrm{par}_G(X) \to X])_{X \in V}$ their induced possibilistic causal network. Then, for any $\omega \in \Omega$ and any family $\Lambda(\omega) = (\alpha_i(\omega))_{i=1}^n$ of correctness degrees α_i that satisfy $0 \leq \alpha_i(\omega) \leq \alpha(\Pi^V_{\mathrm{par}_G(X_i)}(\omega), \Pi^V_{\{X_i\}}(\omega), X_i)$, the quantity

$$\mathrm{Nonspec}\left[\mathcal{N}_G(\mathcal{D})\right]\left(\omega, \Lambda(\omega)\right) = \sum_{i=1}^{n} H\left(\left[\varphi[\mathcal{D}; \mathrm{par}_G(X_i) \to X_i](\omega)\right]_{\alpha_i(\omega)}\right)$$

is called the *nonspecificity of* $\mathcal{N}_G(\mathcal{D})$ *w.r.t.* ω *and* $\Lambda(\omega)$. If we assume uniform distributions on Ω and all $\Lambda(\omega)$, $\omega \in \Omega$, then

$$E\Big(\mathrm{Nonspec}\left[\mathcal{N}_G(\mathcal{D})\right]\Big) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \int_{\Lambda(\omega)} \mathrm{Nonspec}\left[\mathcal{N}_G(\mathcal{D})\right]\left(\omega, \Lambda(\omega)\right) \, \mathrm{d}F_{\Lambda(\omega)}$$

is called the *expected nonspecificity of* $\mathcal{N}_G(\mathcal{D})$.

Algorithm 2.

Input: A database $\mathcal{D}$ of sample cases and a number $k \in \mathbb{N}$.

Output: A DAG $G = (V, E)$ for the induction of a possibilistic causal network $\mathcal{N}_G(\mathcal{D})$ that satisfies $|\mathrm{par}_G(X)| \leq k$ for all $X \in V$ and reflects significant causal dependencies in $\mathcal{D}$.

$V^* := V$; $E := \emptyset$;

while $V^* \neq \emptyset$ **do begin**

select a node $X \in V^*$ and a set $W \subseteq V$ of possible parent nodes ($X \notin W$, $|W| \leq k$) such that the quantity

$$m(W,X) = \frac{1}{|\Omega^W| \cdot |\Omega^{\{X\}}|} \times \sum_{\substack{\omega' \in \Omega^W, \omega'' \in \Omega^{\{X\}}: \\ \alpha(\omega',\omega'',X)>0}} \int_0^{\alpha(\omega',\omega'',X)} \frac{1}{\alpha(\omega',\omega'',X)} H\Big(\big[\pi_{\mathcal{D}}(X \mid W = \omega')\big]_\alpha\Big)\, \mathrm{d}\alpha$$

is minimized without affecting the DAG property of G, when the arcs (Y,X) for all $Y \in W$ are added to G;

$E := E \cup \{(Y,X) \mid Y \in W\}$;

$V^* := V^* \setminus \{X\}$;

end

Note that $m(W,X)$ is in analogy with the corresponding quantity used in Algorithm 1 for the relational setting. It is straightforward to define the modified values $m^*(W,X)$ and $m_*(W,X)$, respectively, and to talk about strong/weak independence in the possibilistic framework. It turns out that weak independence coincides with the property of *non-interactivity* that is well-known in possibility theory (Dubois and Prade, 1988). For more details on an axiomatic and semantic treatment of possibilistic independence, see (Campos *et al.*, 1995).

Theorem 3. *Using the weights* $m(W,X)$ *as defined in Algorithm 2, we have*

$$E\Big(\text{Nonspec}\,[\mathcal{N}_G(\mathcal{D})]\Big) = \sum_{X \in V} m\Big(\text{par}_G(X), X\Big)$$

Theorem 4. *Let* $G = (V,E)$ *be a DAG that is constructed by application of Algorithm 2. Then*

$$E\Big(\text{Nonspec}\,[\mathcal{N}_G(\mathcal{D})]\Big) = \min\Big\{E\Big(\text{Nonspec}\,[\mathcal{N}_{G'}(\mathcal{D})]\Big) \mid G' \in \mathcal{G}_k(V) \text{ and } \text{cons}(G) \subseteq \text{cons}(G')\Big\}$$

The above theorems correspond to Theorems 1 and 2 that we obtained in the more restrictive relational setting. Nevertheless, note that the modified definition of the quantity $m(W,X)$ in Algorithm 2 does not change essentially the tractability of the structure induction problem under consideration. Hence, we obtain that an efficient implementation of Algorithm 2 has a time complexity of $O(mn^{k+1}r^{k+1})$, and the extension of the Greedy search algorithm presented in Section 2 runs efficiently in $O(mn^2r^{k+1})$ time.

Example 3. Consider the database $\mathcal{D} = (D_j)_{j=1}^{8}$ of precise cases shown in Table 6. $\mathcal{D}$ corresponds exactly to the relation R studied in Example 2, but it should be recognized that $\mathcal{D}$ and R have different interpretations: R may be viewed as a single correct set-valued specification of the current object state ω_0, whereas $\mathcal{D}$ is a collection of precise specifications of sample object states $\omega_1, \ldots, \omega_8$, obtained in a set C of contexts $c_1, \ldots, c_8$ that may coincide with the context used for ω_0. Assuming that all of these contexts are equally likely, $\mathcal{D}$ induces an imperfect specification $\Gamma = (\gamma, P_C)$ of ω_0 that reflects the imprecision as well as the uncertainty in our knowledge regarding the value ω_0. The definition of Γ is shown in Table 7.

Tab. 6. A database of precise sample cases.

sample case	X_1	X_2	X_3	X_4	X_5
D_1	1	0	1	1	1
D_2	0	1	0	1	1
D_3	0	1	1	1	0
D_4	0	1	0	0	1
D_5	1	0	1	1	0
D_6	0	1	1	0	1
D_7	0	1	1	1	1
D_8	0	1	0	1	0

Tab. 7. Imperfect specification $\Gamma = (\gamma, P_C)$ of ω_0.

context c_j	weight $P_C(\{c_j\})$	sample state ω_j	specification $\gamma(c_j)$ of sample state ω_j
c_1	0.125	ω_1	{ (1,0,1,1,1) }
c_2	0.125	ω_2	{ (0,1,0,1,1) }
c_3	0.125	ω_3	{ (0,1,1,1,0) }
c_4	0.125	ω_4	{ (0,1,0,0,1) }
c_5	0.125	ω_5	{ (1,0,1,1,0) }
c_6	0.125	ω_6	{ (0,1,1,0,1) }
c_7	0.125	ω_7	{ (0,1,1,1,1) }
c_8	0.125	ω_8	{ (0,1,0,1,0) }

If we are given a dependency hypergraph $H = (V, \mathcal{E})$, for example the one in Fig. 4, we can calculate the induced possibilistic network $\mathcal{N}_H(\mathcal{D}) = (\pi_{\Gamma_E})_{E \in \mathcal{E}}$. The corresponding possibility distributions π_{Γ_E} are listed in Table 8.

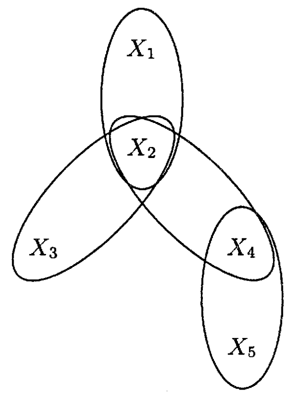

Fig. 4. Dependency hypergraph H.

Tab. 8. Possibility distributions of the possibilistic network $\mathcal{N}_H(\mathcal{D})$.

ω \ E	$\pi_{\Gamma_E}(\omega)$			
	$\{1,2\}$	$\{2,3\}$	$\{2,4\}$	$\{4,5\}$
$(0,0)$	0	0	0	0
$(0,1)$	0.75	0.25	0.25	0.25
$(1,0)$	0.25	0.375	0.25	0.375
$(1,1)$	0	0.375	0.5	0.375

Note that the choice of the dependency hypergraph H is connected with the assumption that H signifies all relevant dependencies among the variables, so that any most specific correct set-valued specification R of ω_0 has a lossless join decomposition $\mathcal{N}_H(R)$, which means that $R = \text{rel}(\mathcal{N}_H(R))$. For this reason, each context c_j can be restricted to a context c_j^E that only describes the frame conditions for specifying the dependencies $\Pi_E^V(\omega_j)$, where $E \in \mathcal{E}$ is an arbitrary hyperedge. Table 9 shows an example for such a restriction, namely the choice of $E = \{1,2\}$, and the resulting imperfect specification Γ_E. Its induced information-compressed representation as the possibility distribution Π_{Γ_E} is characterized in the two left columns of Table 8.

The α-cut $[\pi_{\Gamma_E}]_\alpha$ is the most specific correct set-valued specification of $\Pi_E^V(\omega_0)$ that follows from $\mathcal{D}$, if α-correctness of Γ_E w.r.t. $\Pi_E^V(\omega_0)$ is assumed. As an example, selecting $\alpha = 0.25$ means that at least two of the contexts in C_E do not contradict the context that is used in order to specify $\Pi_E^V(\omega_0)$. In this case, at least two of the set-valued specifications $\gamma_E(c_j^E)$ of $\Pi_E^V(\omega_j)$ are also correct w.r.t. $\Pi_E^V(\omega_0)$, and we obtain $[\pi_{\Gamma_E}]_{0.25} = \{(0,1),(1,0)\}$, $E = \{1,2\}$. If we choose $\alpha = 0.5$, only one of the four candidates $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$ for $\Pi_E^V(\omega_0)$ remains possible, namely $(0,1)$, since we calculate $[\pi_{\Gamma_E}]_{0.5} = \{(0,1)\}$, $E = \{1,2\}$.

Tab. 9. Imperfect specification $\Gamma_E = (\gamma_E, P_E)$ of $\Pi_E^V(\omega_0)$, given $E = \{1, 2\}$.

context	specification	weight
c_j^E	$\gamma_E(c_j^E)$	$P_E(\{c_j^E\})$
c_1^E	$\{(1,0)\}$	0.125
c_2^E	$\{(0,1)\}$	0.125
c_3^E	$\{(0,1)\}$	0.125
c_4^E	$\{(0,1)\}$	0.125
c_5^E	$\{(1,0)\}$	0.125
c_6^E	$\{(0,1)\}$	0.125
c_7^E	$\{(0,1)\}$	0.125
c_8^E	$\{(0,1)\}$	0.125

The selection of any family $\Lambda = (\alpha_E)_{E \in \mathcal{E}}$ of correctness degrees, and the assumption that, for all $E \in \mathcal{E}$, Γ_E is α_E-correct w.r.t. $\Pi_E^V(\omega_0)$, yields most specific correct set-valued specifications $[\pi_{\Gamma_E}]_{\alpha_E}$ if $\Pi_E^V(\omega_0)$ and thus the most specific combined correct set-valued specification $\text{rel}(([\pi_{\Gamma_E}]_{\alpha_E})_{E \in \mathcal{E}}$ of ω_0. An example for such specifications $[\pi_{\Gamma_E}]_{\alpha_E}$ of $\Pi_E^V(\omega_0)$ is shown in Table 10.

Tab. 10. Specifications $[\pi_{\Gamma_E}]_{\alpha_E}$ of $\Pi_E^V(\omega_0)$.

E	α_E	$[\pi_{\Gamma_E}]_{\alpha_E}$
$\{1,2\}$	0.5	$\{(0,1)\}$
$\{2,3\}$	0.3	$\{(1,0),(1,1)\}$
$\{2,4\}$	0.4	$\{(1,1)\}$
$\{4,5\}$	0.3	$\{(1,0),(1,1)\}$

Given the involved correctness assumptions,

$$\omega_0 \in \text{rel}\left([\pi_{\Gamma_E}]_{\alpha_E}\right)_{E \in \mathcal{E}} = \left\{(0,1,0,1,0),(0,1,0,1,1),(0,1,1,1,0),(0,1,1,1,1)\right\}$$

turns out to be the most specific correct set-valued specification of ω_0. Note that the increasing degrees of correctness always lead to the non-decreasing specificity, possibly resulting in a contradictory specification. Selecting, for instance, $\alpha_E = 0.5$ for all $E \in \mathcal{E}$, we calculate $\text{rel}([\pi_{\Gamma_E}]_{\alpha_E})_{E \in \mathcal{E}} = \emptyset$, because $\alpha_E = 0.5$ exceeds the total mass of contexts that can be correct for $\Pi_E^V(\omega_0)$, of $E = \{2,3\}$ or $E = \{4,5\}$.

So far we have seen how to get most specific correct set-valued specifications of ω_0, based on an available possibilistic contraint network $\mathcal{N}_H(\mathcal{D}) = (\pi_{\Gamma_E})_{E \in \mathcal{E}}$ and a family $\Lambda = (\alpha_E)_{E \in \mathcal{E}}$ of assumed correctness degrees of Γ_E w.r.t. $\Pi_E^V(\omega_0)$. From a practical point of view, the problem refers to the fact that at our starting point

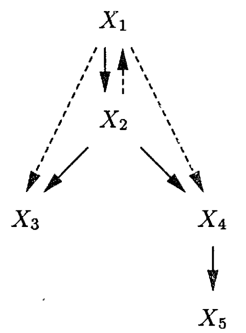

Fig. 5. Scheme of all optimal tree-structured DAGs that can be induced from $\mathcal{D}$.

of considerations, we are normally given the database $\mathcal{D}$, but no information on a dependency hypergraph H and thus an adequate structuring of the data. For this reason, our purpose is to induce H from $\mathcal{D}$. As an approach to a solution to this problem, we presented a method for constructing an optimal DAG $G = (V, E)$ that minimizes the expected amount of information to be provided beyond $\mathcal{D}$ in order to identify or reject any $\omega \in \Omega$ as the current object state with respect to the deductive reasoning scheme induced by G. In this connection, we assumed that there are uniform distributions on Ω and the possible families Λ of correctness degrees. Theorem 3 helps us to find a DAG with the above-mentioned optimality. For this, we need to compute the nonspecificity values $m(W, X)$ for all $X \in V$ and $W \subseteq V \backslash \{X\}$ that satisfy $|W| \leq 1$, where we again restrict ourselves to the class of tree-structured DAGs. Table 11 lists the results of these computations.

Tab. 11. The nonspecificity values $m(W, X)$.

W \ X	X_1	X_2	X_3	X_4	X_5
$\emptyset$	0.667	0.667	0.8	0.667	0.667
$\{X_1\}$	—	0	0.5	0.375	0.875
$\{X_2\}$	0	—	0.5	0.375	0.875
$\{X_3\}$	0.417	0.417	—	0.688	0.792
$\{X_4\}$	0.375	0.375	0.875	—	0.5
$\{X_5\}$	0.708	0.708	0.792	0.417	—

As a consequence of Theorem 3, we find

$$E\Big(\text{Nonspec}\left[\mathcal{N}_G(\mathcal{D})\right]\Big) = 0.667 + 0 + 0.5 + 0.375 + 0.5 = 2.042$$

as the minimum expected nonspecificity of the possibilistic causal network $\mathcal{N}_G(\mathcal{D})$ with a DAG G that is taken from the class $\mathcal{G}_1(V)$ of all tree-structured DAGs w.r.t. V. Figure 5 shows a scheme of all optimal trees.

These trees G and their induced dependency hypergraphs $H(G)$ coincide with those obtained in Section 2, but note that we consider a richer semantic background in the possibilistic setting that leads to more expressive nonspecificity values $m(W, X)$ than in the pure relational setting.

Up to now we confined ourselves to investigate an example of a database with precise cases. But the approach that we introduced in this section together with its background semantics is also appropriate for the treatment of *imprecise cases* and *missing values*. When inducing a structure from a database with *missing values*, the major aim is to find this structure based on dependencies without missing values, and then to complete the rest of the database by postulating the induced structure for all sample cases. If *imprecise cases* are given, we do not intend to complete the database, but rather consider all possible dependencies that may appear as a consequence of knowledge expansion to more specific sample cases.

Referred to our example, suppose to have a simple modification of our database $\mathcal{D}$, saying that

either the value of X_5 in case D_2 has not been observed (a missing value, i.e. $D_2^{(5)}$ is ignored),

or there is no preference regarding the possible values of X_5 in case D_2 (imprecise case, i.e. $D_2^{(5)} = \{0, 1\}$).

In the interpretation as a missing value, given a dependency hypergraph $H = (V, \mathcal{E})$, for all $E \in \mathcal{E}$ that contain X_5, the context c_2^E is regarded as being absent for a specification of the sample dependency $\Pi_E^V(\omega_2)$, and thus C_E changes to $C_E = \{c_1^E, c_3^E, c_4^E, \ldots, c_8^E\}$ with $P_E(\{c\}) = \frac{1}{7}$ for all $c \in C_E$. This modification concerns the nonspecificity values $m(W, X)$ in the way presented in Table 12. Nevertheless, the minimum expected nonspecificity $E(\text{Nonspec}[\mathcal{N}_G(\mathcal{D})])$ remains unchanged, and we get the same set of optimal DAGs.

Tab. 12. Modified nonspecificity values (missing value X_5 in case D_2).

W	$m(W, X_5)$
$\emptyset$	0.875
$\{X_1\}$	0.917
$\{X_2\}$	0.917
$\{X_3\}$	0.917
$\{X_4\}$	0.417
$\{X_5\}$	—

X	$m(\{X_5\}, X)$
X_1	0.708
X_2	0.708
X_3	0.708
X_4	0.5
X_5	—

Turning over to the interpretation as an imprecise sample case, the contexts c_2^E are present, but the context-dependent set-valued specifications of $\Pi_E^V(\omega_2)$ have to be modified as shown in Table 13. The necessary changes in the possibility distributions of the induced possibilistic network $\mathcal{N}_H(\mathcal{D})$ are illustrated in Table 14 and can be compared with Table 8. Note that the occurrence of an imprecise case implies that

Tab. 13. Modified set-valued specifications (imprecision in X_5 of case D_2).

E	$\gamma_E(c_2^E)$
$\{1,5\}$	$\{(0,0),(0,1)\}$
$\{2,5\}$	$\{(1,0),(1,1)\}$
$\{3,5\}$	$\{(0,0),(0,1)\}$
$\{4,5\}$	$\{(1,0),(1,1)\}$

Tab. 14. New definition of $\mathcal{N}_H(\mathcal{D})$.

ω \ E	$\pi_{\Gamma_E}(\omega)$			
	$\{1,2\}$	$\{2,3\}$	$\{2,4\}$	$\{4,5\}$
$(0,0)$	0	0	0	0
$(0,1)$	0.75	0.25	0.25	0.25
$(1,0)$	0.25	0.375	0.25	0.5
$(1,1)$	0	0.375	0.5	0.375

the possibility degrees in at least one of the possibility distributions do no longer sum up to 1.

Again there is an effect of the nonspecificity values $m(W,X)$, as illustrated in Table 15. The set of optimal DAGs does not change, but we now obtain $E(\text{Nonspec}[\mathcal{N}_G(\mathcal{D})]) = 1.98$. Note that a loss of specificity in the database ($D_2^{(5)} = \{0,1\}$ instead of $D_2^{(5)} = \{1\}$) may produce new possible dependencies among variables, which can be weaker or stronger than the old dependencies. In our example, observing a decrease of $E(\text{Nonspec}[\mathcal{N}_G(\mathcal{D})])$, more imprecision has led to a slightly reduced expected nonspecificity of deductive reasoning in the possibilistic causal network $\mathcal{N}_G(\mathcal{D})$, i.e. on average we will need about 2 bits of additional information for identifying or rejecting any $\omega \in \Omega$ as the current object state under consideration.

■

Tab. 15. Resulting changes of nonspecificity values.

W	$m(W,X_5)$
$\emptyset$	0.643
$\{X_1\}$	0.938
$\{X_2\}$	0.938
$\{X_3\}$	0.917
$\{X_4\}$	0.438
$\{X_5\}$	—

X	$m(\{X_5\},X)$
X_1	0.667
X_2	0.667
X_3	0.917
X_4	0.417
X_5	—

4. Application and Concluding Remarks

The main purpose of this paper is to provide concepts for the induction of possibilistic networks from data, based on strong semantics of possibility distributions in an extended random set framework. Considering the presented background, the proposed view of possibilistic reasoning has been implemented in cooperation with Deutsche Aerospace as a part of a project on data fusion problems. We developed a prototype version of the interactive software tool POSSINFER (Gebhardt and Kruse, 1995b; Kruse *et al.*, 1994) for *poss*ibilistic *infer*ence in multidimensional universes of discourse.

POSSINFER runs on SUN workstations under X-Windows and OSF-Motif, and adapts representational features of HUGIN (Andersen *et al.*, 1989) as a powerful tool for knowledge propagation in Bayesian networks. Practical applications confirmed that possibilistic networks have some advantages in the sense that they do not only consider uncertainty, but also imprecision which is typically involved in expert opinions, observations, and measurements. A common effective handling of these two types of imperfect information needs an adequate concept of information compression. Possibilistic networks should therefore be applied only to systems that are not sensitive with respect to the kind of approximate reasoning that has to be tolerated when possibility distributions serve as information-compressed representations of the underlying imperfect specification of a current object state of interest. This is by no means a crucial restriction, since a significant number of industrial applications is in fact not sensitive to approximate modelling. As an example, we mention the impressive success of fuzzy control (Kruse *et al.*, 1994; Zadeh, 1972) as an engineering-related technique of information-compressed interpolation in vague environments (Klawonn *et al.*, 1995).

Referred to the structure induction strategy for possibilistic networks proposed in this paper, we checked our approach with respect to a tutorial example that is discussed in CEC-ESPRIT III Basic Research Project DRUMS II (Defeasible Reasoning and Uncertainty Management Systems). This example deals with the determination of the genotype and verification of the parentage in the F-blood group system of Danish Jersey cattle (Rasmussen, 1992). In the original version, the blood-type determination problem was modelled by using a Bayesian network with the structure shown in Fig. 6, and then implemented under HUGIN for daily use. The induction of the network is supported by 747 sample cases for 9 variables marked as grey nodes. Furthermore, there is additional expert knowledge concerning the relationships between the remaining variables. The expressive power of the example results from the non-trivial, but managable network structure, the consideration of functional, relational, and uncertain dependencies, lots of missing values, and the existence of some interesting exceptional relationships that are only weakly supported by the database.

The most convincing way of handling this example consists in a combined modelling of the two above-mentioned kinds of available information. In this paper, we confined ourselves to structure induction from a database of samples. For this reason, we extended the real database for the 9 marked variables to an artificial database for all 22 variables, while generating the corresponding cases with the aid of the available

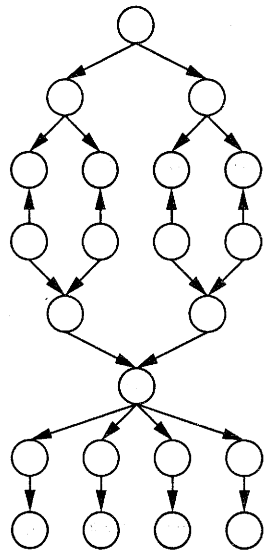

Fig. 6. The blood-type example network.

expert knowledge. Nevertheless, when applying the presented Greedy search modification of Algorithm 2 to the extended database, we could efficiently reconstruct the DAG of Fig. 6 in $O(mn^2r^{k+1})$ time without any erroneous link, except for those dependencies, where a unique directing of arcs is not possible, since it is not expressable in a database. As a simple example consider DAGs G with the same hypergraph $H(G)$ that is induced by their attached moral graphs.

It turned out that the induced possibilistic network, related to its quantitative part, provides propagation results that are, in spite of the underlying information compression, specific enough to obtain a decision quality similar to that of the corresponding Bayesian network. A task of future research is to verify the benefits as well as limits of possibilistic networks in comparison with probabilistic networks from a less pragmatic point of view (Gebhardt and Kruse, 1995a). The first strict theoretical results concerning the calculation of tightest hypertree decompositions of multivariate possibility distributions can be found in (Gebhardt and Kruse, 1996a; 1996b).

Appendix

Proof of Theorem 1

For brevity, let $p(X) \stackrel{\text{def}}{=} \text{par}_G(X)$. Then,

$$E\Big(\text{Nonspec}\,[\mathcal{N}_G(R)]\Big) \stackrel{\text{Def. 2}}{=} \frac{1}{|\Omega|} \sum_{\omega\in\Omega} \sum_{X\in V} \log_2 \left| R\left(X \mid p(X) = \Pi^V_{p(X)}(\omega)\right) \right|$$

$$= \sum_{X \in V} \frac{1}{|\Omega^{p(X)}|} \sum_{\omega \in \Omega^{p(X)}} \log_2 \Big| R\left(X \mid p(X) = \omega\right) \Big|$$

$$\stackrel{\text{Alg. 1}}{=} \sum_{X \in V} m\Big(p(X), X\Big)$$

Proof of Theorem 2

Let $G \in \mathcal{G}_k(V)$ be an output DAG of Algorithm 1. Additionally, let $G' = \mathcal{G}_k(V)$ such that $\text{cons}(G) \subseteq \text{cons}(G')$ is satisfied. For any $G^* \in \mathcal{G}_k(V)$ and any $X \in V$, let

$$\mathcal{W}_{G^*}(X) \stackrel{\text{def}}{=} \Big\{ W \subseteq V \mid |W| \leq k \text{ and } \forall Y \in W : (X, Y) \notin \text{cons}(G^*) \Big\}$$

Then $\text{par}_G(X) \in \mathcal{W}_G(X)$, $\text{par}_{G'}(X) \in \mathcal{W}_{G'}(X)$, $\mathcal{W}_{G'}(X) \subseteq \mathcal{W}_G(X)$, and thus $\text{par}_{G'}(X) \in \mathcal{W}_G(X)$. Hence, due to the minimality of $m(\text{par}_G(X), X)$ relative to $\mathcal{W}_G(X)$, we have

$$m\Big(\text{par}_{G'}(X), X\Big) \geq m\Big(\text{par}_G(X), X\Big)$$

and therefore

$$E\Big(\text{Nonspec}\left[\mathcal{N}_{G'}(R)\right]\Big) \stackrel{\text{Thm. 1}}{=} \sum_{X \in V} m\Big(\text{par}_{G'}(X), X\Big)$$

$$\geq \sum_{X \in V} m\Big(\text{par}_G(X), X\Big)$$

$$\stackrel{\text{Thm. 1}}{=} E\Big(\text{Nonspec}\left[\mathcal{N}_G(R)\right]\Big)$$

The strategy of proving Theorems 3 and 4 is similar to that of Theorems 1 and 2, respectively.

Acknowledgements

We would like to thank S.L. Lauritzen and L.K. Rasmussen for supporting us with the blood-type determination example. Furthermore, we are grateful to C. Borgelt for his helpful comments on a draft version of this paper.

References

Andersen S.K., Olesen K.G., Jensen F.V. and Jensen F. (1989): *HUGIN — A shell for building Bayesian belief universes for expert systems.* — Proc. 11th Int. Joint Conf. *Artificial Intelligence*, pp.1080–1085.

Berge C. (1976): *Graphs and Hypergraphs.* — New York: North Holland.

Buntine W. (1994): *Operations for learning graphical models.* — J. Artificial Intelligence Research, v.2, pp.159–224.

de Campos L.M., Gebhardt J. and Kruse R. (1995): *Axiomatic treatment of possibilistic independence*, In: Symbolic and Quantitative Approaches to Reasoning and Uncertainty (Froidevaux C. and Kohlas J., Eds.), Lecture Notes in Artificial Intelligence, v.946, pp.77–88, Berlin: Springer.

Chow C.K. and Liu C.N. (1968): *Approximating discrete probability distributions with dependency trees.* — IEEE Trans. Information Theory, v.14, pp.462–467.

Cooper G. and Herskovits E. (1992): *A Bayesian method for the induction of probabilistic networks from data.* — Machine Learning, v.9, No.4, pp.309–347.

Dechter R. (1990): *Decomposing a relation into a tree of binary relations.* — J. Comput. Syst. Sci., v.41, No.1, pp.2–24.

Dechter R. and Pearl J. (1988): *Network-based heuristics for constraint satisfaction problems.* — Artificial Intelligence, v.34, No.1, pp.1–38.

Dechter R. and Pearl J. (1989): *Tree clustering for constraint networks.* — Artificial Intelligence, v.38, No.3, pp.353–366.

Dechter R. and Pearl J. (1992): *Structure identification in relational data.* — Artificial Intelligence, v.58, No.2, pp.237–270.

Dempster A.P. (1967): *Upper and lower probabilities induced by a multivalued mapping.* — Ann. Math. Stat, v.38, pp.325–339.

Dempster A.P. (1968): *Upper and lower probabilities generated by a random closed interval.* — Ann. Math. Stat, v.39, pp.957–966.

Dubois D. and Prade H. (1988): *Possibility Theory.* — New York: Plenum Press.

Dubois D. and Prade H. (1991): *Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions.* — Fuzzy Sets and Systems, v.40, No.1, pp.143–202.

Dubois D, Moral S. and Prade H. (1993): *A semantics for possibility theory based on likelihoods.* — CEC–ESPRIT III BRA 6156 DRUMS II, Annual Report.

Gebhardt J. and Kruse R. (1993a): *The context model—An integrating view of vagueness and uncertainty.* — Int. J. Approximate Reasoning, v.9, No.9, pp.283–314.

Gebhardt J. and Kruse R. (1993b): *A comparative discussion of combination rules in numerical settings.* — CEC–ESPRIT III BRA 6156 DRUMS II, Annual Report.

Gebhardt J. and Kruse R. (1993c): *A new approach to semantic aspects of possibilistic reasoning*, In: Symbolic and Quantitative Approaches to Reasoning and Uncertainty (Clarke M., Kruse R. and Moral S., Eds.). — Lecture Notes in Computer Science, v.747, pp.151–160, Berlin: Springer.

Gebhardt J. and Kruse R. (1994): *On an information compression view of possibility theory.* — Proc. 3rd IEEE Int. Conf. *Fuzzy Systems*, pp.1285–1288, Orlando.

Gebhardt J. and Kruse R. (1995a): *Reasoning and learning in probabilistic and possibilistic networks: An overview*, In: Machine Learning: ECML95, (Lavrac N. and Wrobel S., Eds.). — Lecture Notes in Artificial Intelligene, v.912, pp.3–16, Berlin: Springer.

Gebhardt J. and Kruse R. (1995b): *POSSINFER — A software tool for possibilistic inference*, In: Fuzzy Set Methods in Information Engineering: A Guided Tour of Applications (Dubois D., Prade H. and Yager R.R., Eds.). — Chichester: Wiley, (to appear).

Gebhardt J. and Kruse R. (1996a): *Possibilistic graphical models.* — Technical Report, University of Braunschweig, Germany.

Gebhardt J. and Kruse R. (1996b): *Tightest hypertree decompositions of multivariate possibility distributions.* — Proc. Int. Conf. *Information Processing and Management of Uncertainty in Knowledge–Based Systems (IPMU'96)*, Granada, Spain, pp.923–927.

Gyssens M., Jeavons P.G. and Cohen D.A. (1994): *Decomposing constraint satisfaction problems using database techniques.* — Artificial Intelligence, v.66, No.1, pp.57–89.

Hartley R.V.L. (1928): *Transmission of information.* — The Bell Systems Technical J., v.7, pp.535–563.

Hestir K., Nguyen H.T. and Rogers G.S. (1991): *A random set formalism for evidential reasoning*, In: Conditional Logic in Expert Systems, (Goodman I.R., Gupta M.M., Nguyen H.T. and Rogers G.S., Eds.). — New York: North Holland, pp.209–344.

Kampé de Fériet J. (1982): *Interpretation of membership functions of fuzzy sets in terms of plausibility and belief*, In: Fuzzy Information and Decision Processes (Gupta M.M. and Sanchez E., Eds.). — New York: North–Holland, pp.13–98.

Klawonn F., Gebhardt J. and Kruse R. (1995): *Fuzzy control on the basis of equality relations with an example from idle speed control.* — IEEE Trans. Fuzzy Systems, v.3, No.3, pp.336–350.

Kruse R., Gebhardt J. and Klawonn F. (1994): *Foundations of Fuzzy Systems.* — Chichester: Wiley.

Kullback S. and Leibler R.A. (1951): *Information and sufficiency.* — Ann. Math. Statistics, v.22, No.1, pp.79–86.

Kyburg H.E. (1987): *Bayesian and non-Bayesian evidential updating.* — Artificial Intelligence, v.31, No.3, pp.271–293.

Lauritzen S.L. and Spiegelhalter D.J. (1988): *Local computations with probabilities on graphical structures and their application to expert systems.* — J. Royal Stat. Soc., Series B, v.50, No.2, pp.157–224.

Lauritzen S.L, Thiesson B. and Spiegelhalter D. (1993): *Diagnostic systems created by model selection methods—A case study.* — 4th Int. Workshop *AI and Statistics*, Ft. Lauderdale, FL, pp.93–105.

Nguyen H.T. (1978): *On random sets and belief functions.* — J. Mathematical Analysis and Applications, v.65, pp.531–542.

Maier D. (1983): *The Theory of Relational Databases.* — Rockville MD: Computer Science Press.

Pearl J. (1988): *Probabilistic Reasoning in Intelligent Systems — Networks of Plausible Inference.* — San Mateo: Morgan Kaufman.

Pearl J. and Wermuth N. (1993): *When can association graphs admit a causal interpretation (First Report).* — 4th Int. Workshop *AI and Statistics*, Ft. Lauderdale, FL, pp.141–150.

Rasmussen L.K. (1992): *Blood group determination of Danish Jersey cattle in the F-blood group system.* — Dina Research Report 8, Dina Foulum, 8830 Tjele, Denmark.

Saffiotti A. and Umkehrer E. (1991): *PULCINELLA: A general tool for propagating uncertainty in valuation networks*, In: D'Ambrosio B., Smete P. and Bonisonne P.P. (Eds.). — Proc. 7th Conf. *Uncertainty in Artificial Intelligence*, San Mateo: Morgan Kaufman, pp.323–331.

Shafer G. (1976): *A Mathematical Theory of Evidence.* — Princeton: Princeton University Press.

Shafer G. and Pearl J. (1990): *Readings in Uncertain Reasoning.* — San Mateo CA: Morgan Kaufman.

Shenoy P.P. (1989): *A valuation-based language for expert systems.* — Int. J. Approximate Reasoning, v.3, No.5, pp.383–411.

Shenoy P.P. (1993): *Valuation networks and conditional independence.* — Proc. 9th Conf. *Uncertainty in AI*, San Mateo: Morgan Kaufman, pp.191–199,

Shenoy P.P. and Shafer G.R. (1990): *Axioms for probability and belief-function propagation*, In: Uncertainty in AI (Shachter R.D., Levitt T.S., Kanal L.N. and Lemmer J.F., Eds.). — Amsterdam: North Holland, v.4, pp.169–198.

Shore J.E. and Johnson R.W. (1980): *Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy.* — IEEE Trans. Information Theory, v.26, No.1, pp.26–37.

Singh M. and Valtorta M. (1993): *An algorithm for the construction of Bayesian network structures from data.* — Proc. 9th. Conf. *Uncertainty in AI*, pp.259–265, Washington.

Smets P. and Kennes R. (1994): *The transferable belief model.* — Artificial Intelligence, v.66, No.2, pp.191–234.

Spirtes P. and Glymour C. (1991): *An algorithm for fast recovery of sparse causal graphs.* — Social Science Computing Review, v.9, No.1, pp.62–72.

Spohn W. (1990): *A General Non–Probabilistic Theory of Inductive Reasoning*, In: Uncertainty in Artificial Intelligence (Shachter R.D., Levitt T.S., Kanal L.N. and Lemmer J.F, Eds.). — Amsterdam: North Holland, pp.149–158.

Strassen V. (1964): *Meßfehler und Information.* — Zeitschrift Wahrscheinlichkeitstheorie und verwandte Gebiete, v.2, pp.273–305.

Ullman J.D. (1988): *Principles of Database and Knowledge-Base Systems, v.1.* — Rockville, Maryland: Computer Science Press Inc.

Ullman J.D. (1989): *Principles of Database and Knowledge-Base Systems, v.2.* — Rockville, Maryland: Computer Science Press Inc.

Verma T. and Pearl J. (1992): *An Algorithm for deciding whether a set of observed independencies has a causal explanation.* — Proc. 8th Conf. *Uncertainty in AI*, pp.323–330.

Walley P. and Fine T. (1982): *Towards a frequentistic theory of upper and lower probability.* — The Annals of Statistics, v.10, No.3, pp.741–761.

Walley P. (1991): *Statistical Reasoning with Imprecise Probabilities.* — London: Chapman and Hall.

Wang P.Z. (1983): *From the fuzzy statistics to the falling random subsets*, In: Advances in Fuzzy Sets (Wang P.P., Ed.). — Possibility and Applications, New York: Plenum Press, pp.81–96,

Whittaker J. (1990): *Graphical Models in Applied Multivariate Statistics.* — Chichester: Wiley.

Zadeh L.A. (1972): *A rationale for fuzzy control.* — J. Dynamic Systems, Measurement and Control, Series 6, v.94, pp.3–4.

Zadeh L.A. (1978): *Fuzzy sets as a basis for a theory of possibility.* — Fuzzy Sets and Systems, v.1, pp.3–28.