Learning Possibilistic Networks from Data *

Jörg Gebhardt and Rudolf Kruse

Department of Mathematics and Computer Science
University of Braunschweig
D-38106 Braunschweig, Germany
Email: gebhardt@ibr.cs.tu-bs.de

Abstract. We introduce a method for inducing the structure of (causal) *possibilistic* networks from databases of sample cases. In comparison to the construction of Bayesian belief networks, the proposed framework has some advantages, namely the explicit consideration of *imprecise (set-valued) data*, and the realization of a controlled form of *information compression* in order to increase the efficiency of the learning strategy as well as approximate reasoning using local propagation techniques.
Our learning method has been applied to reconstruct a non-singly connected network of 22 nodes and 24 arcs without the need of any a priori supplied node ordering.

1 Introduction

Bayesian networks provide a well-founded normative framework for knowledge representation and reasoning with *uncertain*, but *precise* data. Extending pure probabilistic settings to the treatment of *imprecise (set-valued)* information usually restricts the computational tractability of the corresponding inference mechanisms. It is therefore near at hand to consider alternative uncertainty calculi that provide a justified form of *information compression* in order to support efficient reasoning in the presence of imprecise and uncertain data without affecting the expressive power and correctness of decision making.

Such a modelling approach is appropriate for systems that accept *approximate* instead of crisp reasoning due to a non-significant sensitivity concerning slight changes of information. Possibility theory [21, 5] seems to be a promising framework for this purpose.

In this paper we focus our interest on the concept of a *possibilistic causal network*, which is a directed acyclic graph (DAG) and a family of (conditional) possibility distributions. Since the covering of all aspects of possibilistic reasoning is beyond the scope of this paper, we will confine to the problem of inducing the structure of a possibilistic causal network from data.

In Section 2 we introduce a possibilistic interpretation of databases of set-valued samples. Based on this semantic background, Sections 3 and 4 deal with possibilistic networks and the structure induction method, respectively. In Section 5, we mention some basic ideas and important results including an example of the successful application of our approach.

* This work has partially been funded by CEC-ESPRIT III Basic Research Project 6156 (DRUMS II)

2 Possibilistic Interpretation of Sample Databases

Let $\mathrm{Obj}(X_1, \ldots, X_n)$ be an *object type* of interest, which is 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 dependencies among the variables is available in 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 a 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 nonempty 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 (here named as *contexts*, and denoted by c_j), in which the sample states ω_j have been observed, may only for some of the cases coincide with the context on which the observation of ω_0 is based. The 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* $\Gamma = (\gamma, P_C)$, $C = \{c_1, \ldots, c_m\}$, $\gamma : C \to \mathfrak{P}(\Omega)$, $\gamma(c_j) = D_j$, $j = 1, \ldots, m$, 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 in the same way representative and thus equally likely, then P_C should be the uniform distribution on C.

We suppose that C can be formalized as a subset of a Boolean algebra of 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 uniquely determines 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 [19], [4], and [12]. 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 [7, 8].

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 the position to fully describe the contexts c_j in the form of propositions that are taken from an appropriate underlying Boolean algebra of propositions. For this reason it is convenient to carry out an 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. ω_0, 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

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

Note that, although the application of a probability measure P_C suggests disjoint contexts, we do not make any assumptions about the interrelation of contexts. With respect to the frame conditions that cause the observations, the contexts may be identical, partially corresponding, or disjoint. We add their weights, because disjoint contexts are the "worst case" in which we can not 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*

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

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 follows:

$$\begin{aligned} &[\pi]_\alpha = \{\omega \in \Omega \mid \pi(\omega) \geq \alpha\}, \qquad 0 < \alpha \leq 1, \\ &[\pi]_0 = \Omega. \end{aligned}$$

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, $j = 1, \ldots, m$, 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 $\omega_0 = \omega$.

Note that in the recent years several proposals for the semantics of a 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 [21], the axiomatic view of possibility theory using possibility measures [5], one-point coverages of random sets [14], contour functions of consonant belief functions [17], falling shadows in set-valued statistics [20], Spohn's theory of epistemic states [18], and possibility theory based on likelihoods [6]. Ignoring the underlying interpretation of contexts, π_Γ formally coincides with the 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 better be strongly oriented at the concept of α-correctness. For an extensive presentation of this background of possibility theory, we refer to [8, 9]. Special aspects of possibility measures for decision making in this framework have been considered in [10].

3 The Concept of a Possibilistic Network

For the following investigations, we suppose that all relevant dependencies among the variables in V can be represented in a qualitative way with the aid of a *dependency hypergraph* $H = (V, E)$ [2].

From a quantitative point of view, a dependency hypergraph $H = (V, \mathcal{E})$, when applied to a relation $R \subseteq \Omega$, 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. In this connection, let $\Omega^{\{v\}}$ be the domain of the variable $v \in V$. If $W \subseteq V$ is an arbitrary subset of variables, then

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

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. More particularly, defining

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

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

$$R \subseteq \mathrm{rel}(\mathcal{N}_H(R)).$$

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

$$R = \mathrm{rel}(\mathcal{N}_H(R))$$

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

Given a dependency hypergraph $H = (V, \mathcal{E})$, 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 a sample database $\mathcal{D}$, 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 imperfectly specify ω_0. 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., $\gamma_E(c_j^E) \stackrel{\text{def}}{=} D_j^E$.

Under the assumption that the contexts c_j^E in C^E are equally likely, we choose $P_E(c_j^E) \stackrel{\text{def}}{=} \frac{1}{m}$ 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 to the constraint network $\mathcal{N} = (R_E)_{E\in\mathcal{E}}$ as a lossless join decomposition of $\text{rel}(\mathcal{N})$, which is the resulting most specific correct set-valued specification of ω_0.

The following definition introduces an alternative view of a possibilistic network, which incorporates causality aspects in the way that it deals with a directed acyclic graph of qualitative dependencies, and a family of (conditional) possibility distributions.

Definition 1. Let $V = \{X_1, \ldots, X_n\}$ denote 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 (set-valued) 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

$$\begin{aligned}
&\varphi[\mathcal{D}; W \to V] : \ \Omega \to \text{POSS}(\Omega^{\{X\}}),\\
&\quad \varphi[\mathcal{D}; W \to V](\omega) \stackrel{\text{def}}{=} \pi_{\mathcal{D}}(X \mid W = \omega), \text{ where}\\
&\pi_{\mathcal{D}}(X \mid W = \omega)(\omega') \stackrel{\text{def}}{=} P_C\left(\left\{c \in C \mid \omega \in \Pi_W^V(\gamma(c)) \text{ and } \omega' \in \Pi_{\{X\}}^V(\gamma(c))\right\}\right).
\end{aligned}$$

For any $X \in V$, let

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

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

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

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 $\Pi^V_{\{X\}}(\omega_0)$ that follows from $\mathcal{D}$, given the instantiation of the variables in W ($\Pi^V_W(\omega_0) = \omega$) and the α-correctness of $\Gamma_{W\cup\{X\}}$ w.r.t. $\Pi^V_{W\cup\{X\}}(\omega_0)$.

The possibilistic causal network $\mathcal{N}_G(D)$ is an information-compressed representation of our knowledge about ω_0, given the database $\mathcal{D}$ and a DAG $G = (V, E)$ of causal dependencies among the variables in V. The DAG G induces the dependency hypergraph

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

and, incorporating the database $\mathcal{D}$, the possibilistic constraint network $\mathcal{N}_{H(G)}(D)$.

4 Inducing Possibilistic Networks from Data

In this section we present a new method for inducing a possibilistic causal network $\mathcal{N}_G(D)$ and therefore its attached possibilistic constraint network $\mathcal{N}_{H(G)}(D)$ from a database $\mathcal{D}$ of sample cases. This method is based on viewing a DAG as an abstract deductive reasoning scheme. More particularly, 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 $v_1 < v_2 \cdots < v_n$ reflects this ordering. For any $\omega \in \Omega$, we either want to identify ω as the current object state ($\omega = \omega_0$) or to verify $\omega \neq \omega_0$ by marking those variables v_i, $i = 1, \ldots, n$, for which $\Pi^V_{\{v_i\}}(\omega)$ turns out to be inconsistent with the available general knowledge about ω_0. This knowledge is encoded by $\mathcal{N}_G(D)$ and 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 $\Pi^V_Z(\omega_0)$ within the current object state ω_0. Let $\omega' = \Pi^V_{\mathrm{par}_G(X)}(\omega)$ and $\omega'' = \Pi^V_{\{X\}}(\omega)$. 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

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

is the maximum possible correctness degree of Γ_Z w.r.t. $\Pi^V_Z(\omega)$. 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 $\pi_i := \pi_{\mathcal{D}}(X \mid \mathrm{par}_G(X) = \Pi^V_{\mathrm{par}_G(X)}(\omega))$;
  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 $\Pi^V_Z(\omega_0)$; note that Γ_Z is α_Z-correct w.r.t. $\Pi^V_Z(\omega_0)$;
    provide further information such that either $\Pi^V_{\{X\}}(\omega_0)$ is identified within the set $\Phi_i := [\pi_i]_{\alpha_Z}$ of all the 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 erraneous instantiation has been detected $(\omega \neq \omega_0)$

end

end

The amount of information that has to be added in order to identify $\Pi^V_{\{X\}}(\omega_0)$ within the set Φ_i of possible alternatives can be quantified by $H(\Phi_i)$, where H denotes the *Hartley measure of information* [11].

For any nonempty 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 $H(\emptyset) \stackrel{\text{def}}{=} 0$.

Based on Hartley information, we measure the nonspecificity of $\mathcal{N}_G(D)$ with respect to $\omega \in \Omega$, namely the total amount of additional information (beyond D) 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 ($\omega \neq \omega_0$). We assume that all choices of α_Z-correctness degrees in the interval $[0, \alpha(\omega', \omega'', X)]$ are equally likely, and that in our a priori state of knowledge, we are indifferent concerning the value of ω_0. Summarizing these ideas, we obtain the following definition.

Definition 3. Let $G = (V, E)$ be a DAG, $\mathcal{D}$ a database of sample cases, and $\mathcal{N}_G(\mathcal{D}) = (\varphi[\mathcal{D}; \mathrm{par}_G(X) \rightarrow X])_{X \in V}$ 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 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

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

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

Assuming uniform distributions on Ω and all $\Lambda(\omega)$, $\omega \in \Omega$, then

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

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

The concept of the expected nonspecificity of a causal constraint network is helpful for inducing an optimal DAG G that minimizes $E(\mathrm{Nonspec}[\mathcal{N}_G(D)])$ relative to $\mathcal{D}$ and a chosen class of DAGs. The following Theorem shows another representation of this quantity, which is more convenient for computational aspects.

Theorem 4. *For any $X \in V$ and any $W \subseteq V$, $X \notin W$, define*

$$m(W, X) \stackrel{\text{def}}{=} \frac{1}{|\Omega^{W}| \cdot |\Omega^{\{X\}}|} \cdot \sum_{\substack{\omega' \in \Omega^{W}, \omega'' \in \Omega^{\{X\}}: \\ \alpha(\omega', \omega'', X) > 0}} \int_{0}^{\alpha(\omega', \omega'', X)} \frac{1}{\alpha(\omega', \omega'', X)} H\left([\pi_{\mathcal{D}}(X \mid W = \omega')]_{\alpha}\right) d\alpha.$$

Then,

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

5 Results and Concluding Remarks

- Let $\mathcal{G}_k(V)$, $k \geq 2$, denote the class of all directed acyclic graphs w.r.t. V that satisfy the condition $|\text{par}(X)| \leq k - 1$ for all $X \in V$. Based on Theorem 4, we developed an *Algorithm G1* for determining a DAG $G \in \mathcal{G}_k(V)$ that minimizes $E(\text{Nonspec}[\mathcal{N}_G(D)])$ among all DAGs in $\mathcal{G}_k(V)$ that satisfy the node ordering constraints of G. Algorithm G1 has a time complexity of $O(mn^k r^k)$, where r is defined as $r = \max\{|\Omega^{(i)}| | i = 1, \ldots, n\}$. It does not need any presupposed node ordering, and although it is only optimal w.r.t. a subclass of $\mathcal{G}_k(V)$, it nevertheless tends to deliver a good choice w.r.t. $\mathcal{G}_k(V)$.
- Eleminating arcs of a DAG $G \in \mathcal{G}_k(V)$ in order to get a more simple DAG $G' \in \mathcal{G}_{k-1}(V)$, is connected with a loss of information which is quantified by the corresponding increasement of $E(\text{Nonspec}[\mathcal{N}_{G'}(D)])$ in comparison to the expectation $E(\text{Nonspec}[\mathcal{N}_G(D)])$.
- Algorithm G1 can be made more efficient with the aid of a Greedy search method, starting with $\mathcal{G}_2(V)$ and stepwise extending the optimal output graph with respect to those arcs that reflect the strongest causal dependencies (i.e. the smallest degree of nonspecificity) to the class $\mathcal{G}_k(V)$. The resulting *Algorithm G2* has a time complexity of $O(mn^2 r^k)$.
- From a graph theoretical point of view, independence in the DAG of a possibilistic causal network is represented in the same way as independence in Bayesian networks. Due to the different uncertainty calculus, independence in our approach turns out to basically coincide with the concept of *non-interactivity* well-known from possibility theory. Non-interactivity satisfies the basic properties of independence as proposed in [15], with the exception of the intersection axiom.
- From database theory it is well-known that given a relation R and any hypergraph H, deciding whether $\text{rel}(\mathcal{N}_H(R)) = R$ is NP-hard. Constructing a lossless join decomposition of a relation within a class of dependency hypergraphs is presumably intractable even in cases where each individual member of the class is tractable [3]. Since the problem of structure identification in relational data can be viewed as a special case of the corresponding problem in the possibilistic framework, it is out of reach to get a general, non-heuristic, efficient algorithm for possibilistic learning. One of the tasks of our future research in this field therefore consists in working out, how far Algorithms G1 and G2 deliver tight approximations of optimal (i.e. maximum specificity preserving) decompositions.

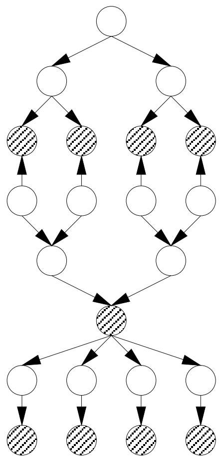

Fig.1 : The blood–type example network

– Algorithms G1 and G2 have successfully been applied for reconstructing the network shown in Figure 1. The underlying application refers to a Bayesian approach, implemented with HUGIN [1] for daily use in Denmark. It deals with the determination of the genotype and verifying the parentage in the F-blood group system of Danish Jersey Cattle [16]. The application is supported by a real database of 747 sample cases for the 9 attributes marked as grey nodes in Figure 1, including lots of missing values. There is also additional expert knowledge regarding the quantitative dependencies among other attributes. Using this information we extended the database to an artificial database for all attributes. Running Algorithm G2 on this database, the network could be efficiently reconstructed in the possibilistic setting without erraneous links, except from those dependencies, where a unique directing of arcs is not possible, since not expressable in a database. We will include our learning strategy in the system POSSINFER [13], a software tool for *poss*ibilistic *infer*ence that we develop in cooperation with Deutsche Aerospace in the field of data fusion problems.

Acknowledgements

We would like to thank S.L. Lauritzen and L.K. Rasmussen for supporting us with the bloodtype determination example.

References

1. S.K. Andersen, K.G. Olesen, F.V. Jensen, and F. Jensen. HUGIN — A Shell for Building Bayesian Belief Universes for Expert Systems, *Proc. 11th Int. Joint Conf. on AI*, 1080–1085, 1989.
2. C. Berge. *Graphs and Hypergraphs.* North Holland, New York, 1976.
3. R. Dechter and J. Pearl. Structure Identification in Relational Data, *Artificial Intelligence*, 58:237–270, 1992.

4. A.P. Dempster. Upper and Lower Probabilities Generated by a Random Closed Interval, *Ann. Math. Stat.*, 39:957–966, 1968.
5. D. Dubois and H. Prade. *Possibility Theory.* Plenum Press, New York, 1988.
6. D. Dubois, S. Moral, and H. Prade. A Semantics for Possibility Theory Based on Likelihoods, CEC–ESPRIT III BRA 6156 DRUMS II, Annual Report, 1993.
7. J. Gebhardt and R. Kruse. The Context Model — An Integrating View of Vagueness and Uncertainty, *Int. J. of Approximate Reasoning*, 9:283–314, 1993.
8. J. Gebhardt and R. Kruse. A Comparative Discussion of Combination Rules in Numerical Settings, CEC–ESPRIT III BRA 6156 DRUMS II, Annual Report, 1993.
9. J. Gebhardt and R. Kruse. A New Approach to Semantic Aspects of Possibilistic Reasoning, In: M. Clarke, R. Kruse, and S. Moral, editors. *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, Lecture Notes in Computer Science, 747, 151–160, Springer, Berlin, 1993.
10. J. Gebhardt and R. Kruse. On an Information Compression View of Possibility Theory, *Proc. 3rd IEEE Int. Conf. on Fuzzy Systems*, Orlando, 1994.
11. R.V.L. Hartley. Transmission of Information, *The Bell Systems Technical J.*, 7:535–563, 1928.
12. J. Kampé de Fériet. Interpretation of Membership Functions of Fuzzy Sets in Terms of Plausibility and Belief. In: M.M. Gupta and E. Sanchez, editors. *Fuzzy Information and Decision Processes*, North–Holland, 13–98, 1982.
13. R. Kruse, J. Gebhardt, and F. Klawonn. *Foundations of Fuzzy Systems*, Wiley, Chichester, 1994.
14. H.T. Nguyen. On Random Sets and Belief Functions, *J. of Mathematical Analysis and Applications*, 65:531–542, 1978.
15. J. Pearl. *Probabilistic Reasoning in Intelligent Systems — Networks of Plausible Inference.* Morgan Kaufman, San Mateo 1988.
16. L.K. Rasmussen. Blood Group Determination of Danish Jersey Cattle in the F-blood Group System, *Dina Research Report 8*, Dina Foulum, 8830 Tjele, Denmark, November 1992.
17. G. Shafer. *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.
18. W. Spohn. A General Non–Probabilistic Theory of Inductive Reasoning, In: R.D. Shachter, T.S. Levitt, L.N. Kanal. and J.F Lemmer, editors, *Uncertainty in Artificial Intelligence*, 149–158, North Holland, Amsterdam, 1990.
19. V. Strassen. Meßfehler und Information, *Zeitschrift Wahrscheinlichkeitstheorie und verwandte Gebiete*, 2:273–305, 1964.
20. P.Z. Wang. From the Fuzzy Statistics to the Falling Random Subsets, In: P.P. Wang, editor. *Advances in Fuzzy Sets, Possibility and Applications*, 81–96, Plenum Press, New York, 1983.
21. L.A. Zadeh. Fuzzy Sets as a Basis for a Theory of Possibility, *Fuzzy Sets and Systems*, 1:3–28, 1978.