

**Part I:**  
**Invited Papers**



# Reasoning and Learning in Probabilistic and Possibilistic Networks: An Overview <sup>\*</sup>

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.** Graphical modelling is a powerful framework for reasoning under uncertainty. We give an overview on the semantical background and relevant properties of probabilistic and possibilistic networks, respectively, and consider knowledge representation and independence as well as evidence propagation and learning such networks from data. Whereas Bayesian networks and Markov networks are well-known for a couple of years, we also outline the perspectives of possibilistic networks as a tool for the efficient information-compressed treatment of uncertain *and* imprecise knowledge.

## 1 Introduction

Due to its simplification of multivariate data analysis, graphical modelling is being increasingly recognized as an appropriate framework for both knowledge representation and inference under uncertainty [65]. Applications can be found in all areas of analysing dependent observations, for example, in regression analysis, spatial analysis, and expert systems. As a consequence of its primary origin in multivariate statistics, the most advanced numerical approaches to the structured handling of uncertain information have been obtained in the field of probabilistic graphical models [4].

Referred to uncertain reasoning in knowledge-based systems, *Bayesian networks* are established as a powerful tool [40]. They provide a well-founded normative framework in the presence of *uncertain*, but *precise* data. On the other hand, extending pure probabilistic settings to the treatment of *imprecise (multi-valued)* information usually restricts the computational tractability of inference mechanisms. Since the explicit modelling of imprecise data is more and more claimed as being necessary for industrial practice, it is reasonable to investigate graphical models related to alternative uncertainty calculi. Under the aspect of efficiency, such uncertainty calculi should provide a justified form of *information compression* and *simplification* in order to support reasoning under uncertainty

---

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

and imprecision without essentially affecting the expressive power and correctness of decision making procedures.

*Possibility theory* [14] seems to be very promising for this purpose. It is a good choice for systems that accept *approximate* instead of crisp reasoning, characterized by a non-significant sensitivity concerning slight changes of information. From this it follows that possibility theory may grow up to play the same role in the field of uncertain reasoning in knowledge-based systems as nowadays *fuzzy control* as a tool for (information-compressed) interpolation between crisp points in vague environments in the field of control engineering [33].

Related to the mentioned perspectives, this paper gives a brief comparison of probabilistic and possibilistic networks: Section 2 addresses the semantic background and relevant properties of probabilistic networks, and summarizes how to perform inference in such networks. Furthermore, the problem of inducing probabilistic networks from a database of sample cases is considered. Section 3 deals with the same topics referred to *possibilistic networks*. Some additional remarks in Section 4 conclude the discussion.

## 2 Probabilistic Networks

Graphical models provide a unified qualitative framework for representing and reasoning with uncertainty and independencies. The qualitative part of a graphical model is a graphical representation of dependencies between variables, expressed, for example, by a directed acyclic graph (DAG), an undirected graph (UG), or a chain graph. Each structure reflects a certain kind of independence in the way that conditioning on evidence (obtained through instantiation of a subset of variables) can be realized by efficient local propagation algorithms, related, for example, to the vertices involved in a hyperedge of a dependency hypergraph, rather than by inefficient global conditioning procedures. For this reason, graphical models are always referred to a notion of (*conditional*) *independence* that has to be specified for the particular uncertainty calculus under consideration [7, 40].

Let  $V$  denote a finite set of variables taking their values  $\omega_v$  on attached domains  $\Omega_v$ ,  $v \in V$ . Furthermore, let  $\Omega$  denote their joint domain, and  $\Omega_A$  the marginal domain of all variables in  $A \subseteq V$ . Finally, let  $\omega_A = (\omega_v)_{v \in A}$  denote a possible state in  $\Omega_A$ , and  $\omega = (\omega_v)_{v \in V}$  a possible state in  $\Omega$ .

In this paper, we refer to finite domains only, which means that we confine to the important case of *discrete graphical models* with  $V$  being the underlying set of vertices.

In pure probabilistic approaches, uncertain knowledge about the quantitative dependencies between the variables in  $V$  are formalized with the aid of a probability distribution  $P$  on  $\Omega$ . Conditional independence, encoded by the qualitative part of a probabilistic graphical model, is defined as follows:

Let  $A$ ,  $B$ , and  $C$  be three disjoint subsets of variables in  $V$ .

$A$  is called *conditionally independent of  $B$  given  $C$  w.r.t.  $P$* , abbr.  $I_P(A, B|C)$ , iff

$$P(\omega_{A \cup B} | \omega_C) = P(\omega_A | \omega_C) \cdot P(\omega_B | \omega_C) \quad (1)$$

whenever  $P(\omega_C) > 0$ ,  $\omega \in \Omega$ .

Equation (1) reads as follows: The marginal conditional probability on  $\Omega_{A \cup B}$ , given any instantiation of the variables in  $C$ , equals the product of the marginal conditional probabilities on  $\Omega_A$  and  $\Omega_B$ , respectively, referred to the same instantiation of the variables in  $C$ .

Given an *independence model*  $M$ , that is, a set of conditional independence statements about the variables in  $V$ , the task is to find a graphical representation  $G$  such that the topology of  $G$  reflects the properties of  $M$  [40, 56].

A probabilistic graphical model can be equated with the set of probability distributions on  $\Omega$  that satisfy the constraints implied by  $G$ . Two probabilistic graphical models are equivalent, if their underlying sets of satisfying probability distributions are equivalent, i.e. they constitute the same joint probability distribution  $P$  on  $\Omega$ .

Among the various approaches to probabilistic graphical models, for example, influence diagrams used to represent decision processes [54, 44, 26], we restrict ourselves to the special case of *Bayesian networks* and *Markov networks*, respectively, which are an advanced and widely discussed framework for knowledge representation and propagation in probabilistic expert systems.

**Bayesian Networks** The most popular kind of a probabilistic graphical model in artificial intelligence is the *Bayesian network*, also called *belief network* [40]. A Bayesian network consists of a directed acyclic graph  $G = (V, E)$ , and a set of conditional probability distributions  $P_v(\omega_v | \omega_{par(v)})$ ,  $v \in V$ , where  $par(v) = \{v' \in V | (v', v) \in E\}$  denotes the set of all parents of  $v$  in  $G$ .

Since  $G$  is directed, Bayesian networks are appropriate to represent direct causal dependencies between variables, which in many cases is quite natural for the purpose of knowledge representation in probabilistic expert systems that support, for example, diagnostic reasoning (abductive inference) in medical applications.

Bayesian networks specify a conditional decomposition of the joint probability distribution  $P$  on  $\Omega$ :

A strictly positive probability distribution  $P$  on  $\Omega$  *factorizes* w.r.t. a directed acyclic graph  $G = (V, E)$ , if

$$P(\omega) = \prod_{v \in V} P(\omega_v | \omega_{par(v)}). \quad (2)$$

In this case,  $P$  satisfies the *global Markov property*, saying that, for any disjoint subsets  $A$ ,  $B$ , and  $C$  of  $V$ , such that  $C$  separates  $A$  from  $B$  in the moral graph induced by  $A$ ,  $B$ , and  $C$ , we have  $I_P(A, B | C)$  [34].

In this sense, a Bayesian network is also a graphical representation of a Markov chain. Alternatively, the Markov property can be expressed in terms of the concept of *d-separation* [62].

It can be shown that  $P(\omega) = \prod_{v \in V} P_v(\omega_v | \omega_{par(v)})$  constitutes the factorizing probability distribution of a Bayesian network, where  $P(\omega_v | \omega_{par(v)})$  equals  $P_v(\omega_v | \omega_{par(v)})$ , whenever  $P_v(\omega_{par(v)}) > 0$ . For this reason, equation (2) can be seen as the *interpretation* of a Bayesian network.

**Markov Networks** An alternative form of probabilistic graphical models refers to undirected graphs  $G = (V, E)$  and is called *Markov network* [40]. It represents Markov random fields, used, for instance, in imaging and spatial reasoning [2] and stochastic models in neural networks [27].

Similar to Bayesian networks, the aim is to get a decomposition of the joint probability, but now in terms of a *potential representation* of  $P$ :

A strictly positive probability distribution  $P$  on  $\Omega$  *factorizes* according to an undirected graph  $G = (V, E)$ , if

$$P(\omega) = \prod_{C \in \text{cliques}(G)} \phi_C(\omega), \quad (3)$$

with  $\text{cliques}(G)$  denoting the set of *maximal cliques* on  $G$ . A maximal clique on  $G$  is a subgraph of  $G$  that is fully connected, but not strictly contained in other fully connected subgraphs of  $G$ . The function  $\phi_C$  depends on  $\omega$  through  $\omega_C$  only.

It can be shown that  $P$  factorizes, if and only if any variable  $v \in V$  is independent of  $V - \{v\} - \text{neighbours}(v)$ , where  $\text{neighbours}(v) = \{v' \mid \{v, v'\} \in E\}$  [24]. The latter condition is called the *local  $G$ -Markovian property* [18].

Similar to DAGs, factorization also implies the *global Markovian property*, saying that, for any disjoint subsets  $A$ ,  $B$ , and  $C$  of  $V$ , such that  $C$  separates  $A$  from  $B$  in  $G$ , we have  $I_P(A, B \mid C)$ . Equation (3) can therefore be viewed as the *interpretation* of a Markov network.

## 2.1 Reasoning in Probabilistic Networks

DAGs and UGs represent conditional independence relations in fundamentally different ways. In particular, there are UGs that represent a conditional independence relation that cannot be represented by a single DAG, and vice versa. On the other hand, every dependence model that is isomorphic to a chordal graph  $G$ , which means that every cycle in  $G$  of at least length four has an edge joining two nonconsecutive vertices, is also isomorphic to a DAG.

Chordal graphs constitute the class of *decomposable models* [40], which have the useful property that the cliques of their Markov networks form a tree.

These special structures supported the development of efficient inference engines for probabilistic graphical models. Their main purpose is to provide a computational simple conditioning of  $P$ . Simplicity of computation means that due to the factorization property of the joint probability distribution  $P$  considered in the probabilistic graphical model, conditioning can be performed stepwise on marginal probability distributions referred to only few variables.

Oriented at the way the human mind reasons, J. Pearl [39] developed a local propagation algorithm that works in singly connected Bayesian networks. S. Lauritzen and D. Spiegelhalter [35] approached the same problem from a pure mathematical point of view. Their proposed method works directly and efficiently in an arbitrary sparse Bayesian network. It involves the transformation of the given DAG into a triangulated UG, and the creation of a tree whose

vertices are the cliques of this triangulated graph. To realize the propagation process, probabilities in the original Bayesian network are updated by message passing among the vertices of this tree of cliques.

Flexible software tools for applying the method are, for example, HUGIN [1, 31] and BAIES [6].

An alternative technique for local computation in hypertrees in the more general framework of *valuation-based systems (VBS)* [47] has been proposed in [51, 46] and implemented in PULCINELLA [43].

## 2.2 Learning Probabilistic Networks from Data

For many application fields, the need to specify probabilistic networks is considerable, so that an automatic induction of the qualitative part (the structure) and the quantitative part (the conditional probability distributions) using underlying background knowledge could alleviate such specification difficulties. Some modelling problems arise from the fact that various kinds of a priori information about the network can be available, expert knowledge as well as an additional database of sample cases, both to be considered in a unified framework for realizing the network induction process.

If we restrict ourselves to a pure data-oriented approach, then the typical assumption is that of a given database  $\mathcal{D} = (D_1, \dots, D_m)$  of  $m$  partial observations  $D_j \subseteq \Omega_{A_j}$  of independent random variables with common distribution  $P$ .

The data are *complete*, if  $A_j = V$  for all  $j = 1, \dots, m$ , *partially incomplete* (with missing values), if  $V = A_1 \cup \dots \cup A_m$ , and *incomplete* (with hidden variables), if  $V \neq A_1 \cup \dots \cup A_m$ .

The network induction process consists of constructing a network within a selected class of networks that best fits the database with respect to a chosen (information-theoretical) optimality criterion.

*Quantitative network induction* refers to estimating the joint probability distribution  $P$  from a parametrized family of probability distributions, given a specific network structure.

A lot of research has been done in this field, referred to basic methods such as maximum likelihood, maximum penalized likelihood, or fully Bayesian approaches, involving different computational techniques of probabilistic inference such as the expectation maximization (EM) algorithm, Gibbs sampling, Laplace approximation, and Monte Carlo methods. For an overview, see [3, 55].

*Qualitative network induction* is learning a network structure from a database of sample cases. In very general terms, the corresponding methods can be classified in three groups, either based on linearity and normality assumptions [42], the extensive testing of conditional independence relations (CI tests) [63], or taking a Bayesian approach [5, 36]. The first group is quite restrictive, CI tests tend to be unreliable unless the volume of data is enormous, and they become rapidly computationally infeasible with increasing number of vertices. Bayesian learning requires debatable a priori assumptions (for example, default uniform priors on distributions, uniform priors on DAGs) and also tends to be inefficient unless Greedy search methods are used.

A quite promising forward search Bayesian learning algorithm is *K2* [5], which approximates the most probable structure of a Bayesian network in  $O(mn^2u^2r)$  time, where  $r$  is the maximum cardinality of the considered domains, and  $u$  a presupposed upper bound on the number of parents each vertex may have. *K2* has successfully been applied, but needs a pre-ordering of the vertices. For this reason, hybrid algorithms, combining CI tests (for finding a vertex ordering) and *K2* (for constructing the Bayesian network with respect to this vertex ordering) have been developed [52]. *K2* works with complete and crisp data, i.e.  $|D_j| = 1$ ,  $j = 1, \dots, m$ , whereas the treatment of missing values and hidden variables is only clear from a theoretical point of view [5].

An alternative algorithm, using a backward search strategy, is described in [30].

An important topic, namely determining the sample size needed to accept the result delivered by structure induction algorithms on a chosen statistical significance level has so far not been considered in a sufficient way. Besides the problem of measuring distances between structures, an additional problem is that of evaluating Markov equivalent DAGs, which are indistinguishable from data alone.

### 3 Possibilistic Networks

#### 3.1 Possibility Distributions

A *possibility distribution*  $\pi$  on a universe of discourse  $\Omega$  is a mapping from  $\Omega$  into the unit interval, i.e.  $\pi : \Omega \rightarrow [0, 1]$  [67].

From an intuitive point of view, given a possibility distribution  $\pi$  as an imperfect specification of a state  $\omega_0 \in \Omega$ , for any  $\omega \in \Omega$ ,  $\pi(\omega)$  quantifies the possibility degree of truth of  $\omega = \omega_0$ . The case  $\pi(\omega) = 0$  means that  $\omega = \omega_0$  is impossible,  $\pi(\omega) = 1$  means that  $\omega = \omega_0$  is possible without any restrictions, and  $\pi(\omega) \in (0, 1)$  means that  $\omega = \omega_0$  is partly possible, since there is evidence that supports  $\omega = \omega_0$  as well as evidence that contradicts  $\omega = \omega_0$ .

A possibility distribution  $\pi$  induces a partial ordering on the possible states, which is less informative than a probability distribution on  $\Omega$ , so that the *measure of possibility*  $\Pi(A)$  of a finite event  $A \subseteq \Omega$  is simply defined as the maximum of  $\pi(\omega)$  for all  $\omega \in A$ .

Recent years of research provided different proposals for the semantics of a *theory of possibility* as a framework for reasoning with uncertain and imprecise data. Among the numerical approaches, we like to mention possibility distributions as epistemic interpretation of fuzzy sets [67], the axiomatic view of possibility theory based on the concept of a possibility measure [14], Spohn's theory of epistemic states [58], possibility distributions as one-point coverages of random sets [38, 28], contour functions of consonant belief functions [45], falling shadows in set-valued statistics [64], and possibility theory based on likelihoods [13].

The view of a possibility distribution as an *information-compressed* representation of an uncertain and imprecise specification of a state  $\omega_0$  of interest can be clarified in a random set framework that generalizes traditional approaches given in [59, 12, 9].



Let  $(C, 2^C, P)$ ,  $C = \{c_1, c_2, \dots, c_m\}$ , denote a finite probability space, and  $\gamma : C \rightarrow 2^\Omega$  a set-valued mapping.  $C$  is interpreted as a set of contexts that are distinguished for set-valued specifications of  $\omega_0$ . The contexts are supposed to describe different physical and observation-related frame conditions, which are, for example, formalized by logical propositions.

The relation  $\gamma(c_j)$  is assumed to be the *most specific correct set-valued specification* of  $\omega_0$ , implied by the frame conditions that describe  $c_j$ . This says that  $\omega_0 \in \gamma(c)$  is guaranteed to be true for  $\gamma(c)$ , but for no proper subset of  $\gamma(c)$ . The quantity  $P(\{c_j\})$  is the (subjective) probability of applying  $c_j$  for the specification of  $\omega_0$ . The resulting *random set*  $\Gamma = (\gamma, P)$  is an imperfect (set-valued and uncertain) specification of  $\omega_0$ . Let  $\pi_\Gamma$  denote the *one-point coverage of  $\Gamma$*  (the *possibility distribution induced by  $\Gamma$* ), defined as

$$\pi_\Gamma : \Omega \rightarrow [0, 1], \pi_\Gamma(\omega) = P(\{c \in C \mid \omega \in \gamma(c)\}).$$

In a complete modelling, the contexts in  $C$  must be specified in detail, so that the relationships among all contexts  $c_j$  and their context-dependent specifications  $\gamma(c_j)$  of  $\omega_0$  are clarified. On the other hand, if the contexts are unknown or ignored, then  $\pi_\Gamma(\omega)$  is the total mass of all contexts  $c$  that provide a correct specification  $\gamma(c)$  of  $\omega_0$ , and this quantifies the *possibility of truth* of “ $\omega = \omega_0$ ” [20, 19].

The (information-compressed) imperfect knowledge about  $\omega_0$ , represented by  $\pi_\Gamma$ , can be specialized in the way that uncertainty is excluded by stating  *$\alpha$ -correctness of  $\Gamma$  w.r.t.  $\omega_0$* , which means that there exists a subset  $C' \subseteq C$  of contexts such that  $P(C') \geq \alpha$  and  $\forall c' \in C' : \omega_0 \in \gamma(c')$ . In this case, the  *$\alpha$ -cut*  $[\pi_\Gamma]_\alpha = \{\omega \mid \pi_\Gamma(\omega) \geq \alpha\}$  of the possibility distribution  $\pi_\Gamma$  turns out to be the most specific correct set-valued specification of  $\omega_0$ , given the representation  $\Gamma$  of our background knowledge about  $\omega_0$ , and the  $\alpha$ -correctness assumption w.r.t.  $\omega_0$  [21].

Operating on possibility distributions can be performed within the pure random set background [9, 28], but has the disadvantage that it is not conform with the *extension principle* [66], which from a semantical point of view has been claimed as the adequate way of generalizing operations from crisp or imprecise (set-valued) data to the possibilistic setting [15, 33]. For this reason, it seems to be more convenient to base the interpretation of possibility degrees on the above mentioned concepts of  $\alpha$ -correctness and maximum specificity, strongly oriented at the meaning of the underlying contexts in  $C$ . For an extensive presentation of this view of possibility theory, we refer to [21, 23]. It verifies the extension principle not just as a principle, but as a theorem in the underlying formal and semantical framework. Special aspects of possibility measures for decision making have been considered in [22].

### 3.2 Possibilistic Graphical Models

Possibilistic graphical models can be introduced in a similar way as probabilistic graphical models. The main difference consists in the change from probability

distributions to possibility distributions, so that there is the need for an appropriate concept of *possibilistic independence*.

Although well-known for a couple of years [29], a unique concept of possibilistic independence has not been fixed yet. For some recent discussions, see [16, 17]. In our opinion, the main problem is to recognize that possibility theory is a calculus for uncertain *and* imprecise reasoning.

For comparison, note that using a single *probability distribution* covers only modelling of *uncertain*, but *precise* data. In a corresponding way, relational database theory applies *relations* in order to represent *imprecise*, but *certain* information about dependencies among variables. Due to their restrictions to distinguishable kinds of imperfect data, the two frameworks supply different concepts of independence, which are probabilistic independence and lossless-join decomposability. Probabilistic independence may be viewed as an *uncertainty-driven* type of independence, whereas lossless-join decomposability [37, 60] reflects an *imprecision-driven* type of independence.

Since possibility theory addresses both kinds of imperfect knowledge, concepts of possibilistic independence can be uncertainty-driven or imprecision-driven, so that there are at least two ways of introducing and justifying them.

Concerning the treatment of imperfect information, one should consider two levels of reasoning, namely the *credal level*, where all operations on our pieces of knowledge take place, and the *pignistic level*, where the final step of decision making follows [53]. Imprecision-driven possibilistic independence is strongly oriented at the credal level, applying the extension principle as the basic concept of operating on possibility distributions, and avoiding normalization, which would change their meaning from quantifying absolute to relative degrees of possibility. In opposite to this, an uncertainty-driven approach to possibilistic independence should be referred to the pignistic level, taking decision making aspects into account and thus quantifying the relative degrees of possibility of events. On this level, the need of normalization is obvious.

With respect to this consideration, in [8] two definitions of possibilistic independence have been justified, namely uncertainty-driven possibilistic independence based on *Dempster's rule of conditioning* [45], specialized to possibility measures, and imprecision-driven possibilistic independence that coincides with the well-known concept of *possibilistic non-interactivity* [14], which is a generalization of the lossless-join-decomposition property to the possibilistic setting. Due to its consistency with the extension principle, we confine to possibilistic non-interactivity. As a concept of possibilistic independence it can be formalized as follows:

Let  $A$ ,  $B$ , and  $C$  be three disjoint subsets of variables in  $V$ .

$A$  is called *conditionally independent* of  $B$  given  $C$  w.r.t.  $\pi$ , abbr.  $I_\pi(A, B|C)$ , iff

$$\pi(\omega_{A \cup B} | \omega_C) = \min\{\pi(\omega_A | \omega_C), \pi(\omega_B | \omega_C)\} \quad (4)$$

whenever  $\pi(\omega_C) > 0$ ,  $\omega \in \Omega$ , with  $\pi(\cdot|\cdot)$  denoting the non-normalized conditional possibility distribution, i.e.

$$\pi(\omega_A | \omega_C) = \max\{\pi(\omega') | \omega' \in \Omega, \omega'_A = \omega_A, \omega'_C = \omega_C\}.$$

It has to be emphasized that both mentioned types of possibilistic independence satisfy the so-called *semi-graphoid axioms* [41] of symmetry, decomposition, weak union, and contraction, which were proposed as basic properties that any concept of conditional independence should have [7, 57, 54].

Possibilistic independence based on Dempster’s rule additionally satisfies the intersection axiom and is therefore a *graphoid* [17]. But note that the intersection axiom is related to uncertainty-driven independence as used in probability theory. Hence, relational independence does not satisfy this axiom, and therefore it cannot be satisfied by possibilistic non-interactivity as its more general type of imprecision-driven independence.

Factorization of possibility distributions is similar to factorization of probability distributions by changing the product to the min-operator.

### 3.3 Reasoning in Possibilistic Networks

The axiomatic framework of *valuation-based systems (VBS)* [47, 50] is able to represent various uncertainty calculi such as probability theory, Dempster-Shafer theory, and possibility theory. Conditional independence in VBSs has been defined in [48] and shown to satisfy the graphoid axioms in [49]. Possibilistic independence in VBSs corresponds to uncertainty-driven possibilistic independence based on Dempster’s rule. Hence, using this type of possibilistic independence, a local propagation algorithm for hypertrees, developed in the setting of VBSs [51, 46], can directly be applied to possibilistic networks that have or can be transformed into this structure (note the correspondence between hypertrees, Markov trees, and join trees). The general algorithm has been implemented in the PULCINELLA system [43].

Choosing possibilistic non-interactivity as the appropriate type of possibilistic independence in order to hold consistency with the extension principle, the VBS approach has to be slightly modified, since no normalization takes place. The related local propagation algorithms for hypertree structures are considered in [33] and implemented in the tool POSSINFER [33].

### 3.4 Learning Possibilistic Networks from Data

Inducing the qualitative part (the structure) and the quantitative part (the conditional possibility distributions) of a possibilistic network from a database of sample cases has not been studied in much detail yet. A first result concerning this topic is presented in [23]. It refers to the semantic background of possibility distributions considered in Section 3.1, and defines a possibilistic interpretation of a database  $\mathcal{D}$  of complete, potentially imprecise sample cases, so that, in opposite to the methods mentioned in Section 2.4,  $|D_j| > 1$  is accepted. Based on this interpretation, the task is to find a possibilistic network that best fits  $\mathcal{D}$  relative to a class of possibilistic networks and a chosen optimality valuation. An obstacle for the corresponding structure induction methods is the fact that there are already serious complexity problems in the more special case of structure identification in relational data: If we are given a hypergraph  $H$  and

a relation  $R$ , then only in cases where  $H$  is tractable (for instance, if  $H$  is a hypertree), one can (tractably) decide whether  $H$  induces a lossless join decomposition of  $R$ . On the other hand, the lossless join decomposition of a relation into a structure taken from a *class* of dependency hypergraphs turns out to be a harder problem, which is presumably intractable even in cases where each individual member of the class is tractable [11]. As a consequence, at least from the viewpoint of efficiency, *heuristics* for inducing possibilistic networks from data have to be tolerated.

In our approach, the valuation of a DAG or an UG with respect to a given database is referred to the amount of information that has to be added to the network in order to identify any possible state  $\omega \in \Omega$  as the unknown state  $\omega_0$  of interest, assuming a uniform distribution on  $\Omega$ . Due to the involved handling of imprecise data, the underlying information measure is *Hartley information* [25], generalized to the uncertainty calculus of possibility theory. In [23], we present the theoretical background and a Greedy search algorithm for inducing DAG structures of possibilistic networks in  $O(n^2mr)$  time without any a priori ordering of the vertices. This algorithm has successfully been applied for reconstructing a non-singly connected DAG with 22 vertices and 24 arcs, based on a generated database of 700 samples [23]. The reconstruction is perfect, except from those dependencies, where a unique directing of arcs is not possible, since not expressible in a database.

## 4 Concluding Remarks

In this overview, we discussed various aspects of reasoning and learning related to probabilistic and possibilistic graphical models. Whereas probabilistic networks deal with uncertain, but precise information, possibilistic networks – for reasons of efficiency – accept a well-founded type of information compression, reducing exact reasoning to only approximate reasoning, which has turned out to be sufficient for many purposes in practice.

Current research interest concerns the automatic construction of graphical models from expert knowledge and available databases of sample cases. In more general terms, this is a well-known problem in the field of *Machine Learning*, where several techniques of concept learning from examples have been discussed. On the other hand, the special properties of the considered network structures and uncertainty calculi restrict the applicability of general machine learning methods. Within the addressed numerical settings, *probably approximately correct (PAC) learning* [61] seems to be convincing at least with respect to its underlying probabilistic framework. In PAC learning, there are some nice complexity results [32, 10] that could be of some interest for limiting the number of sample cases needed in a database in order to guarantee a certain approximation quality of the obtained network. Nevertheless, pure statistical methods of learning probability distributions and verifying probabilistic independence seem to be appropriate. In a similar way, the induction of possibilistic networks from data can be oriented at the specific properties of possibility theory.

## 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. In *Proc. 11th international joint conference on artificial intelligence*, pages 1080–1085, 1989.
2. J. Besag, J. York, and A. Mollie. Bayesian image restauration with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, 43(1):1–59, 1991.
3. W. Buntine. Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, 2:159–225, 1994.
4. P. Cheeseman and R.W. Oldford, editors. *Selecting Models from Data*, volume 89 of *Lecture Notes in Statistics*. Springer-Verlag, 1994.
5. G. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.
6. R. Cowell. BAIES – a probabilistic expert system shell with qualitative and quantitative learning. In J. Bernardo, J. Berger, A. Dawid, and A. Smith, editors, *Bayesian Statistics 4*, pages 595–600. Oxford University Press, 1992.
7. A. Dawid. Conditional independence in statistical theory. *SIAM Journal on Computing*, 41:1–31, 1979.
8. L.M. de Campos, J. Gebhardt, and R. Kruse. Syntactic and semantic approaches to possibilistic independence. Technical report, University of Granada and University of Braunschweig, 1995.
9. 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*, pages 13–98. North-Holland, 1982.
10. S. Decatur. Learning in hybrid noise environments: Using statistical queries. In *Proc. 5th Int. Workshop on Artificial Intelligence and Statistics*, pages 175–185, Fort Lauderdale, 1995.
11. R. Dechter and J. Pearl. Structure identification in relational data. *Artificial Intelligence*, 58:237–270, 1992.
12. A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.*, 38:325–339, 1967.
13. D. Dubois, S. Moral, and H. Prade. A semantics for possibility theory based on likelihoods. Annual report, CEC-ESPRIT III BRA 6156 DRUMS II, 1993.
14. D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, New York, 1988.
15. D. Dubois and H. Prade. Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions. *Fuzzy Sets and Systems*, 40:143–202, 1991.
16. L. Farinas del Cerro and A. Herzig. Possibility theory and independence. In *Proc. of the Fifth IPMU Conference*, pages 820–825, 1994.
17. P. Fonck. Conditional independence in possibility theory. In R. López de Mántaras and D. Poole, editors, *Uncertainty in Artificial Intelligence, Proc. of the Tenth Conference*, pages 221–226. Morgan and Kaufmann, 1994.
18. M. Frydenberg. The chain graph markov property. *Scandinavian Journal of Statistics*, 17:333–353, 1990.
19. J. Gebhardt and R. Kruse. A comparative discussion of combination rules in numerical settings. Annual report, CEC-ESPRIT III BRA 6156 DRUMS II, 1993.
20. J. Gebhardt and R. Kruse. The context model — an integrating view of vagueness and uncertainty. *Int. Journal of Approximate Reasoning*, 9:283–314, 1993.
21. 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*, pages 151–160. Springer, Berlin, 1993.
22. J. Gebhardt and R. Kruse. On an information compression view of possibility theory. In *Proc. 3rd IEEE Int. Conf. on Fuzzy Systems*, pages 1285–1288, Orlando, 1994.
  23. J. Gebhardt and R. Kruse. Learning possibilistic networks from data. In *Proc. 5th Int. Workshop on Artificial Intelligence and Statistics*, pages 233–244, Fort Lauderdale, 1995.
  24. D. Geman. Random fields and inverse problems in imaging. In P. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XVIII — 1988, Lecture Notes in Mathematics, 1427*. 1990.
  25. R.V.L. Hartley. Transmission of information. *The Bell Systems Technical Journal*, 7:535–563, 1928.
  26. D. Heckerman. *Probabilistic Similarity Networks*. MIT Press, 1991.
  27. J. Hertz, A. Krogh, and R. Palmer. *Introduction to the Theory of Neural Computation*. Addison Wesley, 1991.
  28. K. Hestir, H.T. Nguyen, and G.S. Rogers. A random set formalism for evidential reasoning. In I.R. Goodman, M.M. Gupta, H.T. Nguyen, and G.S. Rogers, editors, *Conditional Logic in Expert Systems*, pages 209–344. North-Holland, 1991.
  29. E. Hisdal. Conditional possibilities, independence, and noninteraction. *Fuzzy Sets and Systems*, 1:283–297, 1978.
  30. S. Højsgaard and B. Thiesson. BIFROST – block rekursive models induced from relevant knowledge, observations, and statistical techniques. *Computational Statistics and Data Analysis*, 1994.
  31. F.V. Jensen. *drHUGIN — A System for value of information in Bayesian Networks*. 1993.
  32. M. Kearns. *The Computational Complexity of Machine Learning*. MIT Press, 1990.
  33. R. Kruse, J. Gebhardt, and F. Klawonn. *Foundations of Fuzzy Systems*. Wiley, Chichester, 1994. (Translation of the book: *Fuzzy Systeme*, Series: Leitfäden und Monographien der Informatik, Teubner, Stuttgart, 1993).
  34. S. Lauritzen, A. Dawid, B. Larsen, and H.G. Leimer. Independence properties of directed markov fields. *Networks*, 20:491–505, 1990.
  35. S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Stat. Soc., Series B*, 2(50):157–224, 1988.
  36. S.L. Lauritzen, B. Thiesson, and D. Spiegelhalter. Diagnostic systems created by model selection methods — a case study. In *4th Int. Workshop on Artificial Intelligence and Statistics, January 3–6*, pages 93–105, Ft. Lauderdale, FL, 1993.
  37. D. Maier. *The Theory of Relational Databases*. Computer Science Press, Rockville, MD, 1983.
  38. H.T. Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65:531–542, 1978.
  39. J. Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29:241–288, 1986.
  40. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (2nd edition)*. Morgan Kaufmann, New York, 1992.
  41. J. Pearl and A. Paz. Graphoids: A graph based logic for reasoning about relevance relations. In B.D. Boulay et al., editor, *Advances in Artificial Intelligence*, 2, pages 357–363. North Holland, Amsterdam, 1987.

42. J. Pearl and N. Wermuth. When can association graphs admit a causal interpretation (first report). In *Preliminary Papers of the 4th Int. Workshop on Artificial Intelligence and Statistics, January 3-6*, pages 141–150, Ft. Lauderdale, FL, 1993.
43. A. Saffiotti and E. Umkehrer. PULCINELLA: A general tool for propagating uncertainty in valuation networks. In B. D'Ambrosio, P. Smets, and P.P. Bonisonne, editors, *Proc. 7th Conf. on Uncertainty in Artificial Intelligence*, pages 323–331, San Mateo, 1991. Morgan Kaufmann.
44. R. Shachter. An ordered examination of influence diagrams. *Networks*, 20:535–563, 1990.
45. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.
46. G. Shafer and P.P. Shenoy. Local computation in hypertrees. Working paper 201, School of Business, University of Kansas, Lawrence, 1988.
47. P.P. Shenoy. A valuation-based language for expert systems. *Int. Journal of Approximate Reasoning*, 3:383–411, 1989.
48. P.P. Shenoy. Conditional independence in valuation-based systems. Working Paper 236, School of Business, University of Kansas, Lawrence, KS, 1991.
49. P.P. Shenoy. Conditional independence in uncertainty theories. In D. Dubois, M.P. Wellman, B. D'Ambrosio, and P. Smets, editors, *Uncertainty in Artificial Intelligence: Proc. of the 8th Conference*, pages 284–291, San Mateo, CA, 1992. Morgan Kaufmann.
50. P.P. Shenoy. Valuation-based systems: A framework for managing uncertainty in expert systems. In L.A. Zadeh and J. Kacprzyk, editors, *Fuzzy Logic for the Management of Uncertainty*, pages 83–104. Wiley, New York, NY, 1992.
51. P.P. Shenoy and G.R. Shafer. Axioms for probability and belief-function propagation. In R.D. Shachter, T.S. Levitt, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence (4)*, pages 169–198. North-Holland, Amsterdam, 1990.
52. M. Singh and M. Valtorta. An algorithm for the construction of Bayesian network structures from data. In *Proc. 9th. Conf. on Uncertainty in Artificial Intelligence, Washington*, pages 259–265, 1993.
53. P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.
54. J.Q. Smith. Influence diagrams for statistical modeling. *Annals of Statistics*, 17(2):654–672, 1989.
55. D. Spiegelhalter, A. Dawid, S. Lauritzen, and R. Cowell. Bayesian analysis in expert systems. *Statistical Science*, 8(3):219–283, 1993.
56. P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*, volume 81 of *Lecture Notes in Statistics*. Springer, 1993.
57. W. Spohn. Stochastic independence, causal independence, and shieldability. *Journal of Philosophical Logic*, 9:73–99, 1980.
58. 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*, pages 149–158. North Holland, Amsterdam, 1990.
59. V. Strassen. Meßfehler und Information. *Zeitschrift Wahrscheinlichkeitstheorie und verwandte Gebiete*, 2:273–305, 1964.
60. J.D. Ullman. *Principles of Database and Knowledge-Base Systems*, volume 1. Computer Science Press Inc., Rockville, Maryland, 1988.
61. L. Valiant. A theory of the learnable. *Communications of the ACM*, 27:1134–1142, 1984.

62. T. Verma and J. Pearl. Causal networks: Semantics and expressiveness. In R.D. Shachter, T.S. Levitt, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 69–76. North Holland, Amsterdam, 1990.
63. T. Verma and J. Pearl. An algorithm for deciding if a set of observed independencies has a causal explanation. In *Proc. 8th Conf. on Uncertainty in Artificial Intelligence*, pages 323–330, 1992.
64. P.Z. Wang. From the fuzzy statistics to the falling random subsets. In P.P. Wang, editor, *Advances in Fuzzy Sets, Possibility and Applications*, pages 81–96. Plenum Press, New York, 1983.
65. J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley and Sons, 1990.
66. L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, 9:43–80, 1975.
67. L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.