

On a Tool for Reasoning with Mass Distributions

Rudolf Kruse Erhard Schwecke Frank Klawonn

Department of Computer Science
Technical University of Braunschweig
Bültenweg 74/75
W-3300 Braunschweig
Germany

Abstract

We present a framework for reasoning under uncertainty using mass distributions. Our basic concept is the flow of evidence masses from supersets to subsets which is specified by experts. The expert domain is assumed to have various characteristics so that it can be considered to be a product space with one coordinate for each characteristic. Encoding the qualitative knowledge structure in a dependency network (hypertree) it is possible to describe an efficient propagation algorithm providing mass distributions on each characteristic. This algorithm bases on the computation in (small) subspaces of the product space which reduces its complexity in time as well as in space. It is also easy to make use of parallel computer architectures to improve the efficiency of the algorithm.

1 Introduction

The purpose of this paper is to provide a suitable theoretical tool for the treatment of uncertain information which is an important problem in the range of knowledge based systems [Buchanan and Shortliffe, 1984, Kruse *et al.*, 1991, Pearl, 1988]. So subject to our investigations are statements being not simply true or false but with a validity which is a matter of degree. This may be caused by the fact that the statements' truth depends on some underlying random process, that their 'reliability' reflects the subjective estimates of a human decision maker or that the concept of truth itself has to be weakened.

Throughout this paper we will restrict ourselves to the treatment of subjective valuations of evidence which requires the use of belief functions [Gordon and Shortliffe, 1985, Kohlas, 1988, Kruse *et al.*, 1991, Lowrance *et al.*, 1986, Shafer and Logan, 1987, Shafer *et al.*, 1986, Smets, 1988] measuring the credibility of information although our concept of specialization is very general and can be applied to probabilities as well as to possibility measures.

The expert domain is assumed to be a product space $\Omega^M \stackrel{\text{def}}{=} \Omega_1 \times \dots \times \Omega_m$ with m characteristics $X^{(1)} \in \Omega_1, \dots, X^{(m)} \in \Omega_m$ where $\Omega_i (i = 1, \dots, m)$ is a finite

nonempty set. The qualitative dependency structure can be described by a dependency network reflecting dependencies like $X^{(i)}$ is $A^{(i)} \wedge X^{(j)}$ is $A^{(j)} \implies X^{(k)}$ is $A^{(k)}$ in the subspace $\Omega^{\{i,j,k\}} = \Omega_i \times \Omega_j \times \Omega_k$, whereas the quantitative knowledge about the domain is encoded in mass distributions and specialisation matrices quantifying the flow of evidence masses, the concept we prefer to Dempster's rule of conditioning [Shafer, 1976] often applied in the calculus of belief functions.

Section 2 provides a short outline about mass distributions and belief functions. In section 3 we present our main concept: the flow of evidence masses given by a specialisation matrix [Kruse and Schwecke, 1990]. This concept can be applied to revision.

In order to structure the knowledge about the qualitative dependencies we make use of dependency networks in section 4. Section 5 describes the propagation algorithms we use to draw conclusions from the knowledge. This algorithm was implemented on a TI-Explorer making use of the KEE-shell. It was developed in cooperation with Dornier GmbH, Friedrichshafen, Germany. Applications in the field of data fusion are implemented. Some conclusions concerning the capabilities and restrictions of our tool are presented in the last section.

2 Knowledge Representation with Mass Distributions

As mentioned in the introduction belief functions aim to model a human decision maker's subjective valuation of evidence. For this purpose we consider a finite sample space Ω containing the possible atomic events. The decision maker or 'expert' distributes one unit of 'belief' which we can imagine as movable 'evidence mass' among the elements of Ω with respect to his knowledge about the actual situation, attributing greater amounts to the more likely elements. We will describe such distribution by a mapping $m : \Omega \rightarrow [0, 1]$, $\sum_{x \in \Omega} m(x) = 1$, where the subjective probability of some event $A \subseteq \Omega$ is given by $P(A) = \sum_{x \in A} m(x)$. Unfortunately in practice the expert often is unable to make definite statements, i.e. he fails to specify such distribution in detail. He is only able to value complex events $A \subseteq \Omega$ but not to determine the distribution of that amount of belief committed to A onto its elements. In such cases he specifies a mass distribution m which is a mapping from 2^Ω to the unit interval. Ω in this context is named the frame

of discernment.

Definition 1 Each function $m : 2^\Omega \rightarrow [0, 1]$ is called a mass distribution, whenever

- (i) $m(\emptyset) = 0$ and
- (ii) $\sum_{A: A \subseteq \Omega} m(A) = 1$

hold.

The quantity $m(A)$ is understood to be the measure of 'belief' that is committed exactly to A . m is sometimes called a *basic probability assignment* (bpa). The mass $m(A)$ corresponds to the support given to A but not to any strict subset of A . Those sets A with $m(A) > 0$ are called *focal elements*. To obtain the total measure of belief committed to some set A , we have to sum up the quantities $m(B)$ for all $B \subseteq A$.

Definition 2 If m is a mass distribution on Ω , then the function

$$bel_m : 2^\Omega \rightarrow [0, 1], A \mapsto \sum_{B: B \subseteq A} m(B),$$

is called the *belief function induced by m* .

$bel_m(A)$ represents the degree to which the actual evidence supports A , i.e. it measures the *credibility* of A . We are also able to calculate the degree to which the evidence fails to refute A , i.e. the degree to which A remains *plausible*: $pl_m(A) \stackrel{\text{def}}{=} 1 - bel_m(\bar{A}) = 1 - \sum_{B: B \subseteq \bar{A}} m(B) = \sum_{B: A \cap B \neq \emptyset} m(B)$. We have $bel_m(A) \leq pl_m(A)$, and $bel_m(A) + bel_m(\bar{A}) \leq 1$ for all $A \subseteq \Omega$.

An additional possibility to characterize the information borne by a mass distribution is the concept of *commonality functions*. Let m be a mass distribution defined on Ω . The function $Q_m(A) \stackrel{\text{def}}{=} \sum_{B: A \subseteq B} m(B)$ measures the evidence mass that can freely move to every element or subset of A . This function is called *commonality function*. Obviously $Q_m(A) = 0$ indicates that there is no evidence mass 'above' A , i.e. attached to supersets of A .

By setting $m(\Omega) = 1$ and $m(A) = 0$ for all $A \neq \Omega$ we obtain $bel_m(\Omega) = 1$ and $bel_m(A) = 0$ for all $A \neq \Omega$. This belief function represents *total ignorance* about the domain under consideration, it is therefore called the *vacuous belief function*. On the other hand setting $m(\{x_i\}) = p_i$, $x_i \in \Omega = \{x_1, \dots, x_n\}$ and $m(A) = 0$ for all non-elementary sets A leads to a *Bayesian belief function* or, in terms of probability theory, a discrete probability distribution.

We can imagine 'belief' as partially movable evidence mass, where $m(A)$ is that amount of mass which can, in the light of new information, move to every subset of A but not to sets with elements outside of A .

The concept of revision is based on this idea. So the main impact of supplementary knowledge that the 'truth' is in some set $B \subseteq \Omega$ with certainty is, that all elements of $\bar{B}$ become impossible. All masses attached to subsets A of Ω by a mass distribution m after conditioning with respect to the set $B \subseteq \Omega$ are allocated to the sets $A \cap B$. That means in case of $A \subseteq B$ the masses remain where they are and in case of $A \cap B \neq \emptyset$ they move to $A \cap B$ whereas if $A \cap B = \emptyset$ they have

to be neglected. In total a set $C \neq \emptyset$ obtains the mass $\sum_{A: A \cap B = C} m(A)$. The remaining mass $\sum_{A: A \cap B = \emptyset} m(A)$ arises from partial contradictions of m and the knowledge of B . Since the masses should sum up to one, $1 - \sum_{A: A \cap B \neq \emptyset} m(A)$ is distributed proportionally among those sets emerging as focal elements from the revision process. To formalize this we define the *revised mass distribution* $m(\cdot | B) : 2^\Omega \rightarrow [0, 1]$ by setting

$$m(A|B) \stackrel{\text{def}}{=} \begin{cases} \frac{\sum_{C: B \cap C = A} m(C)}{1 - \sum_{C: C \cap B = \emptyset} m(C)}, & \text{if } A \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

It is defined, if $bel_m(\bar{B}) = \sum_{C: C \subseteq \bar{B}} m(C) = \sum_{C: B \cap C = \emptyset} m(C) \neq 1$. Revising m on $B \subseteq \Omega$ yields the belief function $bel_{m(\cdot | B)}(A) = \frac{bel_m(A \cup \bar{B}) - bel_m(\bar{B})}{1 - bel_m(\bar{B})}$, for all $A \subseteq \Omega$ and the plausibility function $pl_{m(\cdot | B)}(A) = \frac{pl_m(A \cap B)}{pl_m(B)}$, for all $A \subseteq \Omega$. This concept is also known as *Dempster's rule of conditioning* [Shafer, 1976].

3 Specialisation

In order to compare different universes of discourse according to G. Shafer [Shafer, 1976] we introduce the notion of a *refinement*.

Definition 3 A set Ω' is a *refinement* of Ω if there is a mapping $\hat{\Pi} : 2^\Omega \rightarrow 2^{\Omega'}$ such that

- (i) $\hat{\Pi}(\{x\}) \neq \emptyset$ for all $x \in \Omega$,
- (ii) $\hat{\Pi}(\{x\}) \cap \hat{\Pi}(\{x'\}) = \emptyset$, if $x \neq x'$,
- (iii) $\bigcup \{ \hat{\Pi}(\{x\}) \mid x \in \Omega \} = \Omega'$ and
- (iv) $\hat{\Pi}(A) = \bigcup \{ \hat{\Pi}(\{x\}) \mid x \in A \}$.

$\hat{\Pi}$ is called a *refinement mapping*. If such a mapping exists, then the sets Ω and Ω' are compatible, where the refined space Ω' is able to carry more information than Ω . In more formal terms Ω is the quotient space of Ω' .

The next problem we have to solve is to decide for every $x \in \Omega$ whether information concerning some set $A' \subseteq \Omega'$ may be of relevance for the valuation of x or not. For this purpose we define the mapping $\hat{\Pi}^* : 2^{\Omega'} \rightarrow 2^\Omega$, $A' \mapsto \{x \in \Omega \mid \hat{\Pi}(x) \cap A' \neq \emptyset\}$ that attaches to each $A' \subseteq \Omega'$ the set $\hat{\Pi}^*(A')$ of those $x \in \Omega$ which have one or more corresponding elements $x' \in \hat{\Pi}(x)$ within A' . $\hat{\Pi}^*(A')$ is called the *outer reduction* of A' . Note that $\hat{\Pi}^*$ essentially is a projection that attaches to each element $x' \in \Omega'$ that element x_0 for which $x' \in \hat{\Pi}(\{x_0\})$ holds.

From our intuitive interpretation of mass distributions and refinements it becomes obvious that in the course of 'projecting' m' onto Ω the evidence mass $m'(A')$ should be assigned to the outer reduction of A' . No further restriction to subsets of $\hat{\Pi}^*(A')$ is reasonable, since for every $x \in \hat{\Pi}^*(A')$, $A' \neq \emptyset$, there is at least one $x' \in \hat{\Pi}(\{x\})$ in A' possibly turning out to be the final 'truth'. $m'(A')$ might have been assigned just in order

to cover this state of the world. So we obtain the mass distribution

$$\hat{\Pi}^*(m') : 2^\Omega \rightarrow [0, 1], \quad A \mapsto \sum_{\substack{A' \subseteq \Omega' : \\ \hat{\Pi}^*(A') = A}} m'(A'),$$

called the *projection* of m' to Ω . Whenever we have two mass distributions m and m' on Ω and Ω' , respectively, where Ω' is a refinement of Ω with $\hat{\Pi}$ as the respective refinement mapping, then in case that $m = \hat{\Pi}^*(m')$ holds we call m' a *refinement* of m .

Next we have to investigate into the transfer of information in the reverse direction. So the formulation of a mass distribution m on Ω in terms of the refined frame of discernment Ω' is required. The definition is obvious, the suitable mass distribution is

$$\hat{\Pi}(m) : 2^{\Omega'} \rightarrow [0, 1], \quad A' \mapsto \begin{cases} m(A), & \text{if } A' = \hat{\Pi}(A), \\ 0 & \text{otherwise} \end{cases}$$

denoted as the *vacuous extension* of m . From the definition it is clear that each vacuous extension of a mass distribution is its refinement. In contrast to the projection which generally means a loss of information, the vacuous extension preserves the information borne by the original mass distribution.

The main topic of this chapter is to define the concept of specialization. The intuitive idea of a specialization is the projection of a conditioning.

Definition 4 Let s, t be two mass distributions defined on Ω . We call s a *specialization* of t ($s \sqsubseteq t$), if and only if there are two mass distributions s' and t' defined on a refinement Ω' of Ω where s' and t' are refinements of s and t , respectively, and an event $E' \subseteq \Omega'$ s.t. $s' = t'(\cdot | E')$.

In addition to the definition the following theorem gives two equivalent characterizations for the specialization relationship. The first one allows to check easily whether $s \sqsubseteq t$ is valid or not. The second one reflects our intuitive idea of floating evidence masses, it describes the flow of the mass $t(A)$ onto the subsets of A .

Theorem 1 Let s, t be two mass distributions on Ω . The following three statements are equivalent

- (i) s is a specialization of t ,
- (ii) $\forall A \subseteq \Omega : (Q_t(A) = 0 \Rightarrow Q_s(A) = 0)$,
- (iii) For every $A \subseteq \Omega$ there are functions $h_A : 2^\Omega \rightarrow [0, 1]$ such that
 - (a) $\sum_{B: B \subseteq \Omega} h_A(B) = t(A)$,
 - (b) $h_A(B) \neq 0 \Rightarrow B \subseteq A$, for all $B \subseteq \Omega$, and
 - (c) $s(B) = \frac{\sum_{A: A \subseteq \Omega} h_A(B)}{1 - \sum_{A: A \subseteq \Omega} h_A(\emptyset)}$ for all $\emptyset \neq B \subseteq \Omega$.

$h_A(B)$ specifies that portion of 'belief' committed to A that in the course of refining t to s' floats to the set B . Condition (iii.a) of Theorem 1 assures that no evidence mass is lost, condition (iii.b) requires that the masses flow only to subsets. Those masses floating to the empty set represent partial contradictions, thus they have to be neglected and the remaining portions have to be normalized as pointed out in condition (iii.c).

In a more technical notation the specialization-relation can be characterized by a matrix $V : 2^\Omega \times 2^\Omega \rightarrow [0, 1]$, such that, roughly spoken, we obtain the more specific mass distribution m' by 'multiplying' the vector m with the matrix V . In the sequel we use square brackets in order to indicate that we conceive the respective functions as vectors or matrices.

Definition 5 Let Ω be the frame of discernment.

- (i) A matrix $V : 2^\Omega \times 2^\Omega \rightarrow [0, 1]$ is called a *specialization matrix*, if and only if
 - (a) $\sum_{B: B \subseteq \Omega} V[A, B] = 1$ for all $A \subseteq \Omega$
 - (b) $B \not\subseteq A \Rightarrow V[A, B] = 0$.
- (ii) Let V be a specialization matrix and let m be a mass distribution on Ω . If

$$c \stackrel{\text{def}}{=} \sum_{A: A \subseteq \Omega} \sum_{B: B \neq \emptyset} m[A] \cdot V[A, B] > 0$$

then the mass distribution $m \odot V$ is defined by

$$(m \odot V)[B] \stackrel{\text{def}}{=} \begin{cases} \frac{1}{c} \cdot \sum_{A: A \subseteq \Omega} m[A] \cdot V[A, B] & \text{if } B \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

for all $B \subseteq \Omega$.

The main difference between the characterization of a specialization by means of a specialization matrix and by means of mass flow functions h_A , $A \subseteq \Omega$, lies in the fact that the later ones assign absolute amounts of masses whereas the former ones fix relative portions.

Theorem 2 Let m and m' be two mass distributions defined on the frame of discernment Ω . We have

$$m' \sqsubseteq m \iff \exists V : m' = m \odot V,$$

where V is a specialization matrix.

The process of *revision*, i.e. the change from a mass distribution m to the revised mass distribution $m(\cdot | E)$, is in the context of knowledge acquisition and reasoning of importance. Revision is a special case of specialization, therefore it can be equivalently described by specialisation matrices.

Recall that revision with respect to the set $E \subseteq \Omega$ means that those masses bound to sets $A \subseteq E$ remain where they are, while those bound to sets A with $A \cap E \neq \emptyset$ float to $E \cap A$. Masses bound to sets A with $A \cap E = \emptyset$ have to be neglected since they represent (partial) contradictions of the conditioning information E and the mass distribution m . The following definition formalizes these ideas.

Definition 6 Let Ω be the frame of discernment and let $E \subseteq \Omega$ be a non-empty set. The *revision matrix* $R(E) : 2^\Omega \times 2^\Omega \rightarrow [0, 1]$ is defined by

$$R(E)[A, B] = \begin{cases} 1 & \text{if } B = A \cap E \\ 0 & \text{otherwise.} \end{cases}$$

We obtain $m(\cdot | E) = m \odot R(E)$.

4 Reasoning in Product Spaces and Dependency Networks

As mentioned in the introduction the universe of discourse Ω^M can be viewed as a product space of different characteristics of the domain. The quantitative knowledge is encoded by mass distributions on Ω^M and also by specialisation matrices. The aim of the knowledge acquisition process is to provide a set of possible mass distributions given by the experts.

Typically, such a set of mass distributions can be viewed as revisions or more generally specialisations of an a priori distribution.

One cannot expect the experts to give information about the whole product space Ω^M . Usually only dependencies of small groups of characteristics $X^{(i)}$ can be described by the experts. So, what we obtain is knowledge about certain subspaces Ω^S where $S \subseteq M$. Thus it is important to know the qualitative structure of the domain which describes the dependencies between characteristics and subspaces.

The first step of the knowledge acquisition process therefore is to elucidate the qualitative structure of the knowledge resulting in a dependency network with characteristics $X^{(i)}$ ($i \in M$) as vertices and dependency relations as edges. An edge between two vertices $X^{(i)}$ and $X^{(j)}$ ($i \neq j$) represents a direct dependency between these two characteristics which is specified by the experts in the subspace $\Omega^{\{i,j\}}$. The dependency network is directed because of the knowledge being given by rules (e.g. if $X^{(i)}$ is $A^{(i)}$ and $X^{(j)}$ is $A^{(j)}$ then $X^{(k)}$ is $A^{(k)}$). We also assume that all characteristics influence each other mutually (at least indirectly). Otherwise the universe of discourse could be split into different domains which could be modelled independently.

To obtain a propagation algorithm which propagates the knowledge about the subspaces through the dependency network it is important that the network is acyclic. Methods of avoiding cycles can be found in [Kruse and Schwecke, 1989, Shafer and Shenoy, 1988]. For a formal definition of a dependency network see e.g. [Kruse *et al.*, 1991].

The dependency networks enable us to make local computations of the validations given by the experts [Shenoy, 1989].

5 The Propagation Algorithm

In this section we describe the way in which the experts' knowledge is encoded via specialisation matrices and how an 'evidence' $E \subseteq \Omega^M$ obtained from absolutely reliable observations $E^{\{i\}} \subseteq \Omega^{\{i\}}$ induces new (revised) mass distributions on Ω^M .

First the dependency structure of the expert domain has to be represented by a dependency network. Then the experts specify mass flows in certain subspaces by specialisation matrices.

For each dependency of the form $X^{(i_1)}, \dots, X^{(i_n)} \rightarrow X^{(j)}$ the experts provide a specialisation matrix which we consider to be an orthogonal extension matrix, that means the matrix specifies the flow of evidence masses from cylinder sets A in $\mathcal{Q}(\Omega^{\{i_1, \dots, i_n\}})$ to sets

$A \times B$ ($B \in 2^{\Omega^{\{j\}}}$), i.e. the specialisation matrix $V : \mathcal{Q}(\Omega^{\{i_1, \dots, i_n, j\}}) \times \mathcal{Q}(\Omega^{\{i_1, \dots, i_n, j\}}) \rightarrow [0, 1]$ can be represented by an orthogonal extension matrix $V' : \mathcal{Q}(\Omega^{\{i_1, \dots, i_n\}}) \times \mathcal{Q}(\Omega^{\{j\}}) \rightarrow [0, 1]$ where $\mathcal{Q}(\Omega^{\{k_1, \dots, k_t\}}) = \{A_1 \times \dots \times A_t \mid A_\nu \subseteq \Omega^{\{k_\nu\}}, \nu = 1, \dots, t\}$ denotes the cylinders in $\Omega^{\{k_1, \dots, k_t\}}$.

The matrix V' describes how to extend mass distributions on $\Omega^{\{i_1, \dots, i_n\}}$ to $\Omega^{\{i_1, \dots, i_n, j\}}$.

The acquisition of the different matrices is done according to the causal consistent ordering of the dependency network. Starting point is the state of total ignorance, where nothing is known but the 'closed world assumption' mentioned before, which is represented by the mass distribution $m_\nu : \mathcal{Q}(\Omega^\emptyset) \rightarrow [0, 1]$; $m_\nu(\{\epsilon\}) = 1$, the only mass distribution on $\Omega^\emptyset$. So in the state of total ignorance, i.e. before any expert has been asked, all mass distributions on Ω^M are considered possible, because the projection of any mass distribution to $\Omega^\emptyset$ yields m_ν .

In addition to the experts' knowledge reliable observations $E_i \subseteq \Omega^{\{i\}}$ ($i \in M$) have to be considered. (If nothing is observed with respect to $\Omega^{\{j\}}$ we may assume $E_j = \Omega^{\{j\}}$ representing the closed world assumption.) This reliable information $E^M \subseteq \Omega^M$ initiates a revision process revising the possible mass distributions given by experts with respect to E^M . Of course, the computation is not performed in the product space Ω^M but in appropriate subspaces. The results are mass distributions on $\Omega^{\{i\}}$ ($i \in M$) which are projections of the possible revised mass distributions on Ω^M .

The basis of the algorithm we present in the sequel is the idea of independent node processors, exchanging messages. For each quantitative dependency of the form $X^{(i_1)}, \dots, X^{(i_n)} \rightarrow X^{(j)}$ of the dependency network there is one processor $X^{\{i_1, \dots, i_n, j\}}$ which stores the respective orthogonal extension matrix, quantifying the dependencies of the characteristics $X^{(i)}, i \in \{i_1, \dots, i_n, j\}$. All node processors are of the same structure. We have

- an input port to receive messages $E^{(j)}$ from the user, concerning his observation of characteristic $X^{(j)}$
- for each predecessor S a pair of input/output ports to receive a message and to send a message
- for each successor U a pair of input/output ports to receive a message and to send a message
- an output port to provide the marginal revised mass distribution for the user.

The structure of root and leaf nodes is somewhat different. For root nodes no facilities for the communication with fathers are required and similarly leaf nodes need not to communicate with sons.

In the sequel we assume the knowledge acquisition process to be completed, i.e. at every node the orthogonal extension matrices (resp. the marginal mass distribution at the root nodes) are available. In the beginning all observations E^d are assumed to be equal to the respective domain, that means there is no observational knowledge, i.e. for initialization process we have $R(\cdot) = \mathbb{I}$. To bring up the system each node has to send its messages to all neighbours. Since the node processors are assumed to

work independently, this can be done in parallel. The local algorithm for every node is as follows.

If the node processor $X^{(i_1, \dots, i_n, j)}$ receives the observation $E^{(j)}$ from the user it has to perform the following activities

- (i) The calculation of the new marginal revised mass distribution.
- (ii) The calculation of the new messages for the father.
- (iii) The calculation of the new messages for son U .

For details see [Kruse *et al.*, 1991]. There is no feedback, the algorithm terminates after a well defined number of steps. The maximum path length a flow of messages has to cover is therefore given by the diameter of the dependency network. So if a dependency network on the basis of belief functions has to be installed, then at first the qualitative structure of the domain has to be determined by the construction of a dependency network, then the dependencies have to be quantified by the orthogonal extension matrices and after that the dependency network is ready for operation.

Besides the algorithm presented before there are several further aspects, which are important for an useful software tool for the representation of uncertain knowledge on the basis of belief function. For the structure and the man-machine-interface of such system there are some requirements it has to meet, if a successful application is aspired.

First the process of knowledge acquisition has to be supported. The expert should be able to determine first of all the qualitative structure of the domain in form of a hypergraph, whose edges represent the dependencies between the corresponding characteristics. In order to avoid cycles a hypertree covering of the hypergraph has to be constructed [Shafer and Shenoy, 1988]. This should be graphically supported. The change from an arbitrary hypergraph to a dependency network by the join of nodes should be performed either by direct user interventions or automatically. Another possibility to assure that a dependency network arises from this first part of the knowledge acquisition process is to provide only those tools for the construction of the dependency graphs (on the screen) which cannot perform any action destroying the dependency network property. The second step the expert has to perform is to determine the domains for the different characteristics. If a node represents a compound characteristic the single domains $\Omega^{(i)}$ should be specified.

After that the expert has to quantify the qualitative dependencies he stated to exist. That means he has to determine the orthogonal extension matrices. Recall that a typical orthogonal extension matrix has the form

$$V[A^{(i_1, \dots, i_n)}, A^{(j)}], \quad A^{(i_1, \dots, i_n)} \in \mathcal{Q}(\Omega^{(i_1, \dots, i_n)}), \\ A^{(j)} \subseteq \Omega^{(j)}$$

If f.e. $\{i_1, \dots, i_n\} = \{1, 2\}$, $j = 3$, $\text{card}(\Omega^{(1)}) = 6$, $\text{card}(\Omega^{(2)}) = 4$ and $\text{card}(\Omega^{(3)}) = 4$ then we have $\text{card}(\mathcal{Q}(\Omega^{(1,2)})) = 2^6 \cdot 2^4 = 2^{10}$ and $\text{card}(\mathcal{Q}(\Omega^{(3)})) = 2^4$ that means $2^4 \cdot 2^{10} = 16 \cdot 1024 = 16,384$ entries to specify this orthogonal extension matrix. It is obvious that

no expert can determine 16,384 numbers, so he has to be powerfully supported. One possibility to do this is to start with a default assignment of V where

$$V[A^{(i_1, \dots, i_n)}, A^{(j)}] = \begin{cases} 1 & \text{if } A^{(j)} = \Omega^{(j)} \\ 0 & \text{otherwise} \end{cases}$$

which represents total ignorance. Only if the expert explicitly specifies changes of these default values they are modified.

Also the user needs support. Firstly he should have a graphical representation of the dependency network, which enables him to bring in his observations, secondly there should be a graphical presentation of the results to facilitate the correct interpretation of the results. The system should not provide directly the mass distribution but the induced belief- and plausibility function, but not of all sets but of those the expert specified to be of interest.

In addition he should be able to ask for arbitrary sets. Moreover it might be useful to defer the normalization to the output of the results and to display the portion of evidence mass, on which the result is based (e.g. 0.3%) to have a measure of the degree of 'pathology' of the actual case under consideration.

The above example shows the limits of this approach. The complexity of these matrices explodes if the size of the involved sets increases, especially if compound characteristics or nodes with multiple fathers are considered. But if the domains are slightly small (e.g. $\Omega^{(1)} = \{\text{friend/enemy}\}$ etc.) and the dependency network is mainly a tree then this approach can yield useful results.

An additional increasing on efficiency can be obtained by omitting the graphical surface in process control applications or by efficient algorithms for the representation of orthogonal extension matrices, making use of the fact that they tend to be sparse.

6 Conclusions

The main objective of this article is to introduce a software tool for handling uncertainty about imprecise data. Note that in contrast to the work on belief function done by Shafer, Shenoy and Smets where the rule of combination plays a central role we use a mass flow concept where the masses (or probabilities) are attached to the imprecise data (subsets of a frame of discernment).

We assume the *qualitative* knowledge (about dependencies) to be encoded in terms of hypertrees [Shafer and Shenoy, 1988]. This allows to construct a propagation algorithm which gains efficiency by using methods similar to the ones considered by [Shenoy, 1989]. Furthermore we assume the *quantitative* expert knowledge to be encoded in cylinder sets of the product space which leads to specialisation matrices with only a few non-zero entries reducing the expenditure of computer memory.

The result of the reasoning process are mass distributions on the sets $\Omega^{(i)}$ describing the knowledge about characteristics $X^{(i)}$ and being projections of mass distributions on the whole product space.

The method has been implemented on a TI Explorer under KEE and can be used f.e. for data fusion problems.

References

- [Buchanan and Shortliffe, 1984] B.G. Buchanan and E.H. Shortliffe. *Rule-Based Expert Systems*. Addison Wesley, Reading, Massachusetts, 1984.
- [Gordon and Shortliffe, 1985] J. Gordon and E.H. Shortliffe. A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space. *Artificial Intelligence*, 26:323 - 357, 1985.
- [Kohlas, 1988] J. Kohlas. Conditional Belief Structures. *Probability in Engineering and Information Science*, 2:415-433, 1988.
- [Kruse and Schwecke, 1989] R. Kruse and E. Schwecke. On the Treatment of Cyclic Dependencies in Causal Networks. In Proceedings of the 3rd IFSA Congress, pages 416-419, Seattle 1989.
- [Kruse and Schwecke, 1990] R. Kruse and E. Schwecke. Specialization - A New Concept for Uncertainty Handling with Belief Functions. *International Journal of General Systems* (to appear) 1990.
- [Kruse et al., 1991] R. Kruse, E. Schwecke, and J. Heinsohn. *Uncertainty and Vagueness in Knowledge Based Systems: Numerical Methods*. Springer, Berlin, 1991.
- [Lowrance et al., 1986] J. D. Lowrance, T. D. Garvey, and T. M. Strat. A Framework for Evidential-Reasoning Systems. In Proceedings of the AAAI, pages 896-903, Philadelphia, 1986.
- [Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*. Morgan Kaufmann, New York, 1988.
- [Shafer, 1976] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, 1976.
- [Shafer and Logan, 1987] G. Shafer and R. Logan. Implementing Dempster's Rule for Hierarchical Evidence. *Artificial Intelligence*, 33:271 - 298, 1987.
- [Shafer and Shenoy, 1988] G. Shafer and P. Shenoy. Local Computations in Hypertrees. Working Paper No 201. School of Business University of Kansas, 1988.
- [Shafer et al., 1986] G. Shafer, P. Shenoy, and K. Mellouli. Propagating Belief Functions in Qualitative Markov Trees. *International Journal of Approximate Reasoning*, 1:349 - 400, 1986.
- [Shenoy, 1989] P. Shenoy. A Valuation-Based Language for Expert Systems, *International Journal of Approximate Reasoning*, 3:383-411, 1989.
- [Smets, 1988] P. Smets. Belief Functions. In: P. Smets, A. Mamdani, D. Dubois, and H. Prade, (eds.) *Non Standard Logics for Automated Reasoning*. Academic Press, London, 1988, 253 - 286.