Cluster-based Specification Techniques in Dempster-Shafer Theory for an Evidential Intelligence Analysis of Multiple Target Tracks

by

JOHAN SCHUBERT

VETENSKAP OCH KONST

KTH

Stockholm 1994

CLUSTER-BASED SPECIFICATION TECHNIQUES IN DEMPSTER-SHAFER THEORY FOR AN EVIDENTIAL INTELLIGENCE ANALYSIS OF MULTIPLE TARGET TRACKS

by

JOHAN SCHUBERT

KTH

Department of Numerical Analysis and Computing Science
Royal Institute of Technology
S-100 44 Stockholm, SWEDEN

TRITA-NA-9410
ISRN KTH/NA/R--94/10--SE
ISSN 0348-2952

ISBN 91-7170-801-4

ABSTRACT

In Intelligence Analysis it is of vital importance to manage uncertainty. Intelligence data is almost always uncertain and incomplete, making it necessary to reason and taking decisions under uncertainty. One way to manage the uncertainty in Intelligence Analysis is Dempster-Shafer Theory. We may call this application of Dempster-Shafer Theory *Evidential Intelligence Analysis*. This thesis contains five results regarding multiple target tracks and intelligence specification in Evidential Intelligence Analysis.

When simultaneously reasoning with evidence about several different events it is necessary to separate the evidence according to event. These events should then be handled independently. However, when propositions of evidences are weakly specified in the sense that it may not be certain to which event they are referring, this may not be directly possible. In the first article of this thesis a criterion for partitioning evidences into subsets representing events is established.

In the second article we will specify each piece of nonspecific evidence by observing changes in cluster and domain conflicts if we move a piece of evidence from one subset to another. A decrease in cluster conflict is interpreted as an evidence indicating that this piece of evidence does not actually belong to the subset where it was placed by the partition. We will find this kind of evidence regarding the relation between each piece of evidence and every subset. When this has been done we can make a partial specification of each piece of evidence.

In the third article we set out to find a posterior probability distribution regarding the number of subsets. We use the idea that each single piece of evidence in a subset supports the existence of that subset. With this we can create a new bpa that is concerned with the question of how many subsets we have. In order to obtain the sought-after posterior domain probability distribution we combine this new bpa with our prior domain probability distribution.

For the case of evidence ordered in a complete directed acyclic graph the fourth article presents a new algorithm with lower computational complexity for Dempster's rule than that of step by step application of Dempster's rule. We are interested in finding the most probable completely specified path through the graph, where transitions are possible only from lower to higher ranked vertices. The path is here a representation for a sequence of states, for instance a sequence of snapshots of a physical object's track.

The fifth article concerns an earlier method for decision making where expected utility intervals are constructed for different choices. When the expected utility interval of one alternative is included in that of another, it is necessary make some assumptions. If there are several different decision makers we might sometimes be interested in having the highest expected utility among the decision makers. We must then also take into account the rational choices we can assume to be made by later decision makers.

Keywords: Belief functions, Dempster-Shafer theory, evidential reasoning, nonspecific evidence, evidence correlation, cluster analysis, directed acyclic graph, computational complexity, decision making.

To my parents.

This work was done at the National Defence Research Establishment (FOA), Stockholm, Sweden.

ACKNOWLEDGEMENT

I wish to express my sincere appreciation to Prof. Stefan Arnborg, my advisor, Royal Institute of Technology, for his support and constructive criticism.

A deep debt of gratitude goes to Dr. Per Svensson, Director of Research, National Defence Research Establishment (FOA), for providing an excellent research environment and for his careful reading and commenting on style of manuscripts.

I am grateful to my colleague Ms. Ulla Bergsten, FOA, for many discussions on Dempster-Shafer theory applications and for coauthoring one of the articles of this thesis.

Thanks is owed to Dr. John D. Lowrance, SRI International, for interesting discussions and much inspiration during his three week visit to FOA in the summer of 1988.

PREFACE

This thesis is divided into four chapters and five appendices containing the articles "On Nonspecific Evidence" [Johan Schubert, *Int. J. Intell. Syst.* **8**(6), 711-725, 1993], "Specifying Nonspecific Evidence" [Johan Schubert, Manuscript], "Finding a Posterior Domain Probability Distribution by Specifying Nonspecific Evidence" [Johan Schubert, Manuscript], "Dempster's Rule for Evidence Ordered in a Complete Directed Acyclic Graph" [Ulla Bergsten and Johan Schubert, *Int. J. Approx. Reasoning* **9**(1), 37-73, 1993], and "On Rho in a Decision-Theoretic Apparatus of Dempster-Shafer Theory" [Johan Schubert, Manuscript]. The first chapter gives an overview of management of uncertainty, the fundamentals of Dempster-Shafer theory and a brief summary of the state of the art in evidential reasoning, especially concerning belief propagation. The objective of the work, decision support in anti-submarine intelligence analysis is the focus of Chapter 2. In Chapter 3 we give a summary of the results of the five articles. Finally, conclusions are drawn (Chapter 4).

CONTENTS

1. INTRODUCTION

This chapter we will serve as a background to the five articles ["On Nonspecific Evidence," *Int. J. Intell. Syst.* **8**(6), 711-725, 1993; "Specifying Nonspecific Evidence," Manuscript; "Finding a Posterior Domain Probability Distribution by Specifying Nonspecific Evidence," Manuscript; "Dempster's Rule for Evidence Ordered in a Complete Directed Acyclic Graph," *Int. J. Approx. Reasoning* **9**(1), 37-73, 1993; and "On Rho in a Decision-Theoretic Apparatus of Dempster-Shafer Theory"]. In Section 1.1 we will discuss management of uncertainty and its knowledge representation, control strategies and different approaches to combination of evidence. In Section 1.2 we will give the fundamentals of Dempster-Shafer theory needed to understand the five articles and in Section 1.3 we will give a brief summary of the state of the art in evidential reasoning.

1.1. Management of Uncertainty

Management of uncertainty concerns the reasoning process of deciding how to act under uncertainty. In uncertainty management, *knowledge representation* and the calculus for *combination of evidence* is of vital importance. For instance, we must ensure that our *knowledge representation* structure does not violate the models assumption about independence between different items of information. The *control strategies* (Cohen [6, 7]) employed may be of equal importance. Even when uncertainty is not explicitly represented, with a good control strategy it may sometimes be possible to choose actions wisely.

When using intelligent systems to solve problems, the systems' reasoning is guided by the control strategies. We must infer these strategies from the domain expert. For instance, a physician may describe his problem-solving strategy as follows (Cohen [6]):

> *"First I take a history, which usually triggers a few possibilities. I try to narrow the differential as much as possible during the history. Next I do a physical exam. I'm looking for signs that help me to narrow the differential further. The physical can also help me refine my suspicions: The history may tell me there's a cardiac problem, but the physical can narrow it to, say, mitral valve prolapse. I avoid tests, especially invasive ones, and rarely perform them except to confirm something I already believe pretty strongly from the history and physical."*

We recognize four kinds of actions: gathering of evidence through history, physical examination, invasive and noninvasive tests. The initial gathering of evidence is done by the study of history. Further gathering of historical evidence and evidence from the physical examination are combined and used to reduce and refine the set of possible hypotheses. From the remaining hypotheses a subset, focus of attention, is selected, the "differential." Finally, tests are made, but here inference is greatly restricted. The only tests that are made are those that can confirm the hypotheses of attention that are strongly believed.

From this example we see how *control strategies* such as selection of *focus of attention*, *control of inference* and *control of action* can be extracted from the domain problem and used to guide the reasoning and evidence gathering processes. The purpose of the control strategies are to achieve the domain goals, e.g. the diagnosing of a patient, in an efficient manner. If the control strategies are implemented declaratively the system may reason about its own control and decide which control strategies are most appropriate in a particular situation. In the medical domain there might for instance be two different control strategies depending on whether a patient is stable or critical. Evidence gathering might, for instance, be more restricted when the patient is critical because of time requirements.

When evidences are uncertain we must also manage the uncertainty by control strategies. This can be done by treating the reduction in uncertainty as any other goal. The domain goals, such as diagnosing a patient, is now complemented with the goal of reducing uncertainty. Any action to gather evidence may now be chosen both for the purpose of reducing uncertainty and for achieving domain goals.

The introduction of uncertain evidence may actually facilitate the construction of control strategies. For instance, attention may be focused on hypotheses where the evidence gathered by additional actions have the best estimated trade-off between cost and reduction in uncertainty. On the other hand, in order to solve domain goals, we may want to focus on the hypotheses with the highest belief. The evidence gathering action is selected to confirm, disconfirm or depending on the outcome, either confirm or disconfirm the focus of attention.

The control strategy decides how to proceed under uncertainty. By control strategies we mean internal functions in the system used to improve the systems computational strategy by choosing focus of attention, which part of a problem to work on, and to control inferences, what methods to use. Most problems can be broken down into subproblems. The *focus of attention* is selected by the control strategy among the different subproblems. The *control of inference* then decides which reasoning methods to apply to the chosen subproblem. If we did not have a focus of attention or had a constantly changing focus of attention there would be no steering of the reasoning process towards a solution of the problem. If on the other hand the focus of attention is difficult to change, the process may lock into a wrong solution. It is the task of the control of inferences to specify which reasoning methods to apply to the subproblem, i.e. how to solve it, and put a limit on the number of inferences. Without a limit on inference we would falter under a confusion of possibilities. If on the other hand the inferences are too restricted we may not reach important possible solutions and we may also make some false conclusions. Finally, a system might also have a *control of action*. An action is here any event, internal or external, that changes the state of the environment. The control of action is an internal function either taking actions internally in a system or requesting external actions taken. It decides how to interact with the environment. For instance, when should we pay the cost of gathering additional evidence.

One additional problem when reasoning under uncertainty is how to decide when to stop taking evidence gathering actions. The strategies must decide the estimated benefit of a further reduction in uncertainty versus the increasing cost of gathering additional evidence.

There are two main competing philosophies for *combination of evidence* in uncertainty management, that of symbolic truth maintenance by assumption-based truth-maintenance system (ATMS) (de Kleer [13]) and that of numeric propagation. In ATMS uncertainty are represented qualitatively, where propositions are true, false, assumed true, assumed false or unknown. These systems use non-monotonic reasoning, changing and retracting assumptions when internal conflicts are detected to restore consistency. We can integrate symbolic truth maintenance and numeric propagation if probabilities are attached to assumptions in an ATMS and its symbolic machinery is used to compute the probabilities that an arbitrary proposition is provable (Laskey and Lehner [16, 17]), i.e. the belief in the proposition is the probability of provability of the proposition (Pearl [23, 24]). This would enhance the value of the control strategy. For instance, the control of inference could avoid inference rules whose antecedents are unlikely and the control of action could search for evidence that will distinguish between uncertain and competing hypothesis.

The integration could also modify the conflict resolution strategy of the assumption-based truth-maintenance system, now being able to hold conflicting hypothesis as long as the conflict is not too large and only when the conflict is too large would we change defaults to restore consistency in the system.

There are two different approaches to numeric propagation, the "probabilistic" approach based on probability theory (mostly Bayesian theory of subjective probability or its generalization, Dempster-Shafer theory of belief functions) and the "possibilistic" approach based on fuzzy sets. In contrast to the probabilistic approach the uncertainty in the possibilistic approach is not one of truth but one of degree of membership. Here a membership function indicates to which degree an element of the universe belongs to the fuzzy set.

In many problems we are faced with judgement-based probabilities that can not be said to be frequencies of an event. Using subjective probabilities, Bayesian theory, for uncertainty management (Pearl [23]) in these cases might seem natural, but the choice is not without its problems. One major problem when building systems based on subjective probability theory is that the conditional probabilities of all cause-effect relations have to be obtained. When those probabilities are not available, approximate methods must be used to complete the model. For this reason only a few systems have been made using subjective probabilities, predominantly in medicine.

For instance, a team of researchers at Aalborg University have developed Hugin (Andersen *et al.* [1]), an expert system shell for probabilistic reasoning. HUGIN uses a causal probabilistic network with conditional probabilities describing uncertain cause-effect relations. It has been used in the medical domain to construct MUNIN: a knowledge-based assistant for electromyography, using a network of several hundred nodes to diagnose muscle and nerve diseases.

The Dempster-Shafer theory of belief functions (Dempster [8], Shafer [26, 28, 29]) is a generalization of the Bayesian theory. It avoids the problem of having to assign non-available prior probabilities and especially all conditional probabilities of cause-effect relations. That is, no assumptions have to be made about not available probabilities. The belief of a proposition is instead drawn from the sum of probabilities for those propositions that are subsets of the proposition in question. When different pieces of evidence are independent their beliefs can be combined with Dempster's rule to compose an overall belief in the proposition.

There are however two other major reasons in favor of Dempster-Shafer theory. These are the representation of ignorance and the lack of requirement to fix the probability of every propositions negation to one minus the probability of the proposition. In complete ignorance between two outcomes a "Bayesian" would be forced to assign probabilities to the two outcomes that sum to one, say an equal probability of $1/2$ to both outcomes. This makes it impossible to differentiate between the completely ignorant case and a case with known $1/2$ probabilities. In Dempster-Shafer theory, on the other hand, no probability is assigned to an outcome when there is no evidence, leading to a zero belief in both outcomes.

The degree to which an outcome is plausible, i.e. one minus the degree of belief in the negation of the outcome, is in the Bayesian case $1/2$ in both outcomes due to the fixing of probabilities, while they are equal to one in Dempster-Shafer theory since there is no available evidence against the outcomes. Thus, the belief and plausibility in Dempster-Shafer theory make up an attractive representation of ignorance, different from the case with known $1/2$ probabilities.

1.2. A Mathematical Theory of Evidence

In this section we briefly explain the fundamentals of evidence theory (also called Dempster-Shafer theory of belief functions). In evidence theory probability is assigned to a proposition by a basic probability assignment. The proposition states that the truth is in a subset A of an exhaustive set of mutually exclusive possibilities, a frame of discernment Θ. It is not required that any probability is assigned to individual elements of A, nor is it required to assign any probability to the complement of A. The remaining probability that is not assigned to A may be assigned to any other subsets of the frame or to the whole frame of discernment itself.

1.2.1. Basic probability assignment

The basic probability assignment is a function from the power set of Θ to [0, 1]

$$\mathrm{m}:\ 2^{\Theta} \to [0,1]$$

whenever

$$m(\varnothing) = 0$$

and

$$\sum_{A \subseteq \Theta} m(A) = 1$$

where $m(A)$ is called a basic probability number, that is the belief committed exactly to A.

1.2.2. Belief, Plausibility and Commonality functions

The total belief of a proposition A is drawn from the sum of probabilities for those propositions that are subsets of the proposition in question and the probability committed exactly to A

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B)$$

where $\mathrm{Bel}(A)$ is the total belief in A and $\mathrm{Bel}(\cdot)$ is a belief function

$$\mathrm{Bel}\colon\ 2^{\Theta} \to [0,1].$$

A subset A of Θ is called a focal element of Bel if the basic probability number for A is non-zero. The union of all focal elements is called the core of Bel. It is possible to recover the basic probability assignment from the belief function by

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \cdot \mathrm{Bel}(B)$$

where A - B means $A \cap B^c$.

In addition to the belief in a proposition A it is also of interest to know how plausible a proposition might be, i.e the degree to which we do not doubt A. The plausibility,

$$\mathrm{Pls}\colon\ 2^{\Theta} \to [0,1]$$

is defined as

$$\mathrm{Pls}(A) = 1 - \mathrm{Bel}(A^c).$$

We can calculate the plausibility directly from the basic probability assignment

$$\mathrm{Pls}(A) = \sum_{B \cap A \neq \varnothing} m(B).$$

From plausibility we may easily go back to belief and basic probability

$$\mathrm{Bel}(A) = 1 - \mathrm{Pls}(A^c)$$

and

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|+1} \cdot \mathrm{Pls}(B^c).$$

Thus, while belief in A measures the total probability certainly committed to A, plausibility measures the total probability that is in or can be moved into A, i.e. $\mathrm{Bel}(A) \leq \mathrm{Pls}(A)$.

Finally, we are interested in all probability that can freely be moved to any point in A, the commonality number of A.

A commonality function,

$$\mathrm{Q}\colon\ 2^{\Theta} \rightarrow [0,1]$$

is defined by

$$\mathrm{Q}(A) = \sum_{B \supseteq A} m(B).$$

We may also calculate the commonality number from belief or plausibility

$$\mathrm{Q}(A) = \sum_{B \subseteq A} (-1)^{|B|} \cdot \mathrm{Bel}(B^c)$$

and

$$\mathrm{Q}(A) = \sum_{B \subseteq A} (-1)^{|B|+1} \cdot \mathrm{Pls}(B).$$

From commonality we may just as easily go back to belief and plausibility

$$\mathrm{Bel}(A) = \sum_{B \subseteq A^c} (-1)^{|B|} \cdot \mathrm{Q}(B)$$

and

$$\mathrm{Pls}(A) = \sum_{\varnothing \neq B \subseteq A} -1^{|B|+1} \cdot \mathrm{Q}(B).$$

1.2.3. Combination of belief functions

If we receive a second item of information concerning the same issue from a different source, the two items can be combined to yield a more informed view. Combining two belief functions is done by calculating the orthogonal combination with Dempster's rule. This is most simply illustrated through the combination of basic probability assignments. Let A_i be a focal element of Bel_1 and let B_j be a focal element of Bel_2. Combining the corresponding basic probability assignments m_1 and m_2 results in a new basic probability assignment $m_1 \oplus m_2$

$$m_1 \oplus m_2(A) = K \cdot \sum_{A_i \cap B_j = A} m_1(A_i) \cdot m_2(B_j)$$

where K is a normalizing constant

$$K = \Big(1 - \sum_{A_i \cap B_j = \varnothing} m_1(A_i) \cdot m_2(B_j)\Big)^{-1}.$$

This normalization is needed since, by definition, no probability mass may be committed to $\varnothing$. The new belief function $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2(\cdot)$ can be calculated by the above formula from $m_1 \oplus m_2(\cdot)$. The combination of two commonality functions is simply done by taking the normalized product of the two commonality numbers

$$\mathrm{Q}_1 \oplus \mathrm{Q}_2(A) = K \cdot \mathrm{Q}_1(A) \cdot \mathrm{Q}_2(A)$$

where

$$K = \Big(\sum_{\varnothing \neq B \subseteq \Theta} (-1)^{|B|+1} \cdot \mathrm{Q}_1(B) \cdot \mathrm{Q}_2(B)\Big)^{-1}.$$

When we wish to combine several belief functions this is simply done by combining the first two and then combine the result with the third and so forth. As an alternative it is possible to extend the combination to n belief functions. For combining the n corresponding basic probability assignments we have

$$m_1 \oplus m_2 \oplus \ldots \oplus m_n = K \cdot \sum_{A_i \cap B_j \cap \ldots \cap C_k = A} m_1(A_i) \cdot m_2(B_j) \cdot \ldots \cdot m_n(C_k)$$

where

$$K = \Big(1 - \sum_{A_i \cap B_j \cap \ldots \cap C_k = \varnothing} m_1(A_i) \cdot m_2(B_j) \cdot \ldots \cdot m_n(C_k)\Big)^{-1}.$$

1.3. Evidential Reasoning

In a review article on the theory and practice of belief functions [28], Shafer argues that "interactive systems seem appropriate to belief functions, since the theory practically requires that the relation between evidence and questions of interest should be unique to each application." This means that belief networks and numerical judgments must be constructed separately for each case.

A number of interactive evidential reasoning systems have been implemented using Dempster-Shafer theory. Two of the better known are the interactive systems Gister (Lowrance *et al.* [20]) and DELIEF (Zarley *et al.* [40]). Gister which was developed at SRI International contains two main subsystems, the Curator and the Analyzer. In the Curator the user constructs the set of possible propositions and their relations. By using the Analyzer he may examine available evidences. Here evidences can be projected one or more time units according to projection relations, changing the proposition of evidence over time. If two or more evidences are projected to the same time they may be combined by a fusion operator. Finally, belief and plausibility are calculated for propositions through an interpretation operator.

In DELIEF, developed at the University of Kansas, the user graphically creates a network of variables describing propositions and between them joint variables describing relations. After having supplied the evidence to variables and joint variables the evidence can be propagated. For reasons of computational complexity the belief functions are combined step by step using local computations. Recently, there has been a new implementation, TRESBEL (Xu [38]), developed at the Université Libre de Bruxelles, based on DELIEF but with an additionally optimized propagation scheme.

1.3.1. Previous work on decision methodologies

To make decisions under uncertainty is somewhat complicated in Dempster-Shafer theory because of the interval representation. In [22] Nguyen and Walker discussed different approaches to decision making with belief functions. They found three different basic models. The first is based on the Choquet integral that yields the expected utility with respect to belief functions;

$$E_F(u) = \int_0^{\infty} F(u > t)dt + \int_{-\infty}^{0} [F(u > t) - 1]\, dt$$

where F is a belief function defined on 2^{Θ} by $F(A) = inf\{P(A) : P \in \mathbf{P}\}$ and $\mathbf{P} = \{P : F \leq P\}$ is a class of probability measures on Θ. This leads to the pessimistic strategy of ranking alternatives by their minimal expected utility.

In the second basic model the decision maker uses some additional information or subjective views. Instead of searching for the alternative that maximizes expected utility the utility function will be supplemented by some new function dependent on the utility and some other parameter corresponding to the additional information or subjective views. An article by Strat [34] is an example of the second basic model.

The third basic model consists of models using the insufficient reason principle or equivalently the maximum entropy principle.

Smets and Kennes [33] have developed a two-level model of credal belief and pignistic probability, called the 'Transferable belief model' (TBM), that belongs to the third category of Nguyen and Walker [23].

On the credal level of this model the reasoning process takes place in the usual manner as within Dempster-Shafer theory. Here beliefs are held by belief functions and combined by Dempster's rule. One addition in this model is the possibility to choose an 'open-world assumption' where belief may be given to $\varnothing$ and where there is no normalization in Dempster's rule.

When a decision must be taken, the belief on the credal level is transformed to a probability at the pignistic level by a "pignistic transformation" based on Laplace's insufficient reasoning principle;

$$BetP(x) = \sum_{x \subseteq A \in \mathfrak{R}} \frac{m(A)}{|A|} = \sum_{A \in \mathfrak{R}} m(A) \cdot \frac{|x \cap A|}{|A|},$$

$$BetP(B) = \sum_{A \in \mathfrak{R}} m(A) \cdot \frac{|B \cap A|}{|A|}$$

where $BetP(\cdot)$ is the pignistic probability we should use to 'bet' with in a utility maximization process. Here $\mathfrak{R}$ is the set of all propositions. It is called the betting frame. If Π is some partition of the frame of discernment Θ then $\mathfrak{R}$, the betting frame, is the boolean algebra of the subsets of Θ that is generated by Π, e.g. if $\Theta = \{\omega_1, \omega_2, \omega_3\}$ and let us assume that $\Pi = \{\{\omega_1, \omega_2\}, \{\omega_3\}\}$ then $\mathfrak{R} = \{\{\omega_1, \omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2, \omega_3\}\}$. The sets A and B in the equation above are elements of $\mathfrak{R}$ and x which is called an atomic element is an element of $\mathfrak{R}$ that is also an element of Π.

One justification of the pignistic transformation within the TBM is that it is the only transformation that can transform belief functions at the credal level to probability functions at the pignistic level, such that two different belief functions, Bel_1 and Bel_2, with different propositions but with equal frames should yield the same pignistic probability for the disjunctive proposition, $BetP_{12}$, regardless of whether we first transform the two belief functions to two pignistic probability functions, $BetP_1$ and $BetP_2$, by the pignistic transformation and then find the pignistic probability for the disjunctive proposition, $BetP_{12}$, from those two probabilities, or if we do it the other way around by first finding the belief function for the disjunctive proposition, Bel_{12}, and then transform this belief function by the pignistic transformation to the sought after pignistic probability, $BetP_{12}$.

It is obvious that the received pignistic probability regarding some proposition

A depends on the organization of the betting frame $\mathfrak{R}$ itself. But regardless of the organization of the betting frame we always have $BetP(A) \geq Bel(A) \quad \forall A \in \mathfrak{R}$.

Another method for decision making has recently been developed by Strat [34]. In this method an expected utility interval is constructed for each choice;

$$[E_*(x),\ E^*(x)]$$

where

$$E_*(x) \triangleq \sum_{A_i \subseteq \Theta} inf(A_i) \cdot m_\Theta(A_i),$$

$$E^*(x) \triangleq \sum_{A_i \subseteq \Theta} sup(A_i) \cdot m_\Theta(A_i).$$

If the intervals of two different propositions do not overlap we have a clear choice. But if they do we must choose a value from each interval to be able to rank them. Let this value be the expected utility

$$E(x) \triangleq E_*(x) + \rho \cdot (E^*(x) - E_*(x))$$

where ρ is defined as the probability that the ambiguity about the utility of every non-singleton focal element will turn out as favorably as possible, i.e. the probability that nature will turn out as favorably as possibly towards us as decision makers. If both interval limits of the utility interval are higher for one alternative than for another, then this one is always preferable regardless of the value of ρ, otherwise our preference will depend on the assumed value of ρ. When we make an assumption about the value of ρ we should not confuse ambiguity with risk. Our risk preference is handled by adopting utilities. In a comparison with Bayesian decision analysis about necessary assumptions, we are not forced to make the assumptions about non-available probabilities necessary in the Bayesian analysis. Instead we only have to make the assumption about ρ.

Other authors have also looked into decision analysis with belief functions. Lesh [18], for instance, used an empirically derived coefficient to interpolate a point value in the evidential interval he called expected evidential belief rather than choosing a discriminating point in the utility interval. This coefficient reflects the decision makers' preference towards ignorance. Shafer [27], on the other hand, focused on goals since these unlike utilities do not change under refinement of the frame of discernment. In his constructive decision theory goals were related to actions by available evidence offering support that certain goals may be achieved by a specific action, δ. The preference of an action is here based on the difference between the expected total weight of goals achieved and precluded by that action;

$$E(u^{+}(\delta)) = \sum_{i=1}^{n} w_i \cdot \mathrm{Bel}(A_i)$$

and

$$E(u^{-}(\delta)) = \sum_{i=1}^{n} w_i \cdot \mathrm{Bel}(A_i^c).$$

Kleyle and de Korvin [14] focus on the course of action without being directly concerned about a goal or the utility of some proposition. Evidences are simply delivering a probability for the proposition that one course of action is "best". Their model for decision making is based on an elimination process utilizing sequentially acquired information. A simple elimination rule is to eliminate, at the kth cycle, any course of action whose plausibility is less than some other course of action's belief. However, to protect against the possibility of an early wrong elimination, the elimination rule may be modified, e.g. stipulating that a course of action must be chosen for elimination in a number of cycles before it is actually eliminated. After a course of action is eliminated from the frame of discernment the belief structure is updated. This is done by assigning zero mass to basic probability assignments supporting those subsets that include the eliminated action, renormalizing all other basic probability assignments and recalculating belief and plausibility. After having applied the elimination rule and a possible updating any new information that may have arrived is brought into the analysis. When only one course of action is remaining this one is deemed to be the preferred one. Kleyle and de Korvin states that the mass function tends to be ever more focused on subsets containing only a few courses of actions as the elimination and updating loop proceeds, eventually almost entirely focused on singletons. The evidential intervals of every course of action is then shrinking swiftly, making it highly unlikely that there will ever be a state where no course of action can be eliminated. However, to preclude this possibility the best course of action may then simply be chosen by the highest belief after a fixed number of loops without elimination.

1.3.2. Previous work on cluster methodologies

Some work has been done combining cluster methodologies (Jain and Dubes [11]) with Dempster-Shafer theory. Lowrance and Garvey [19] suggested that the conflict could be used as a distance measure between bodies of evidence in some clustering algorithm. Lesh [18] suggested the same and pointed to a work on synthesizing knowledge with cluster methods (Chiu and Wong [5]) that might be of interest. Here, samples where partitioned into clusters using a maximum within-cluster nearest-neighbor distance and then iteratively regrouped by minimizing a probabilistic information measure.

Others have pursued solutions where traditional cluster methods were used to partition evidences into subsets and Dempster-Shafer theory was used to determine the meaning of the subset. In a paper on multiple target tracking for automatic tracking systems Rasoulian, Thompson, Kazda and Parra-Loera [25] considered a clustering technique to associate evidences from different sensors that agree on time and space coordinates. Each cluster was here a representative for an object at a particular time and space coordinate. A distance measure between any two sensor reports was derived from the sensors' gaussian distributions. A relational matrix of thresholded distance measures described when any two sensors had detected some object at the same time and space coordinate. A transformed relational matrix satisfying transitivity was used to create an equivalence relation with which sensor reports could be partitioned into clusters. The sensor reports also carried information about object type and if all reports within a cluster were not of the same object type the meaning of the cluster was determined by fusing all evidences within the cluster by Dempster-Shafer theory. Since the sensor reports either confirmed or disconfirmed a single object, the evidences were in the form of simple support functions focused on singletons or their complements, so that they could be combined in linear time with Barnett's method [2]. Thus, Dempster-Shafer theory was used to determine the meaning of clusters but not used in the clustering process itself.

1.3.3. Previous work on belief propagation in hierarchies

There has been some work on generally applicable improvements of the time complexity of Dempster's rule, e.g. [12, 36], reducing the exponential time complexity in the general case from $O(3^{|\Theta|})$ to $O(|\Theta| \cdot 2^{|\Theta|})$. However, most improvements have concerned important special cases. Foremost among these are methods dealing with belief propagation in trees.

In 1985 Gordon and Shortliffe [9] suggested that when evidence supports singletons or disjoint subsets of the frame, a hierarchical network of subsets could be pruned to a hierarchical tree. The assumption is that a strict hierarchy of hypotheses can be defined from some subsets of 2^{Θ} and that a system will only receive information for these subsets. They proposed a method partly based on the work of Barnett [2] for reasoning about hypotheses with hierarchical relationships.

Barnett showed that simple support functions focused on singletons or their complements can be combined with a time complexity, for each considered subset of Θ, that is linear in the size of the frame, $|\Theta|$. In order to obtain linear time complexity, it is assumed that simple support functions with the same foci have already been combined.

Barnett's method can be described as first combining all simple support functions with equal foci and then, for each singleton, combining the resulting simple support functions for and against the singleton. For each singleton, this results in a separable support function with three focal elements, i.e. the singleton, its complement and Θ. Finally, the separable support functions are combined separately for each considered subset of Θ in such a way that a linear time complexity is obtained;

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B) \triangleq \sum_{\varnothing \neq B \subseteq A} K \cdot \sum_{\cap B_i = B} \prod_{1 \leq i \leq n} m_i(B_i)$$

$$= K \cdot \sum_{\varnothing \neq \cap A_i \subseteq A} \prod_{1 \leq i \leq n} m_i(A_i),$$

where $K = (1 - k)^{-1}$ and k is the conflict.

In the summation over $\cap A_i$, $A_i \in \{ \{i\}, \neg\{i\}, \Theta\}$, in the last term there are only two ways the intersection of all A_i's will not be empty. First, if there is only one A_i that is a singleton and all others are either complements of singletons or offer support for the whole frame, then the intersection of all A_i will be that singleton. The intersection will obviously be empty if there is more than one A_i that is a singleton. Secondly, if there are no A_i's at all that are singletons and there is at least one A_i that offers support to the whole frame, then the intersection is a subset of the frame that contains all elements of the frame except those elements where there is an A_i that offers support for the complement of the element's singleton. When all A_i's offer support for complements of singletons this subset will be empty.

The summation can then be divided into three terms. The first term represents the case where exactly one A_i is a singleton. The second term represents the case where no A_i's are singletons. From the second term we subtract the third term, the case where all A_i offer support to complements of singletons. The second and third terms correspond together to the second way the intersection will not be empty;

$$K \cdot \left(\sum_{\cap A_i \subseteq A \mid \exists j . A_i = \begin{cases} \{i\}, j = i \\ \neg\{i\} \wedge \Theta, j \neq i \end{cases}} \prod_{1 \leq i \leq n} m_i(A_i) + \sum_{\cap A_i \subseteq A \mid A_i \neq \{i\}} \prod_{1 \leq i \leq n} m_i(A_i) \right.$$

$$\left. - \sum_{\cap A_i \subseteq A \mid A_i = \neg\{i\}} \prod_{1 \leq i \leq n} m_i(A_i) \right)$$

$$= K \cdot \left(\sum_{q \in A} m_q(\{q\}) \cdot \prod_{i \neq q} (1 - m_i(\{i\})) \right.$$

$$\left. + \left[\sum_{i \notin A} m_i(\neg\{i\}) \right] \cdot \left[\sum_{i \in A} (1 - m_i(\{i\})) \right] - \prod_{1 \leq i \leq n} m_i(\neg\{i\}) \right)$$

$$= K \cdot \left(\left[\prod_{1 \le i \le n} (1 - m_i(\{i\})) \right] \cdot \left[\sum_{i \in A} \frac{m_i(\{i\})}{1 - m_i(\{i\})} \right] \right.$$

$$\left. + \left[\prod_{i \notin A} m_i(\neg\{i\}) \right] \cdot \left[\prod_{i \in A} (1 - m_i(\{i\})) \right] - \prod_{1 \le i \le n} m_i(\neg\{i\}) \right).$$

From the last equality it is evident that belief for each instance can be calculated in $O(n)$ time, where K is calculated in a similar manner. One realizes that Barnett's technique will also work when the simple support functions are focused on subsets or their complements if all subsets considered are disjoint.

Besides the assumption that the domain allows a hierarchical network to be pruned to a hierarchical tree and that a system will only receive information about those subsets of the frame that are in the tree, the method by Gordon and Shortliffe is approximate in that it does not assign belief to subsets that are not in the tree. It is this approximation that changes the time complexity from exponential to linear.

The first step is borrowed from Barnett's method. All evidences with equal foci, confirming and disconfirming, are combined, with the only difference that what Barnett did with simple support functions focused on singletons is done here for all subsets of the frame that are in the tree, T. The resulting basic probability assignments are calculated for each subset by the simple formula

$$\forall i|\ A_i \wedge A_i^c \in T.m_i(A_i) = 1 - \prod_{\forall j|\ A_j = A_i} (1 - m_j(A_i)) .$$

Now there are two bpa's for each subset of the frame that is in the tree, one confirming the subset and one disconfirming it, we want to combine all bpa's in the entire tree. However, combining bpa's where some focal elements are complements of subsets in the tree might produce an intersection that is not a subset or a complement of a subset that is in the tree. We begin with the confirming bpa's. These are easily combined since the intersection between two focal elements is either empty or the smaller of the two sets. This is because of the tree structure where the focal element of a child is a subset of the focal element of the parent and where focal elements at different branches are disjoint. That is, when m_i is combined with m_j and A_j is above A_i in the tree the intersection of A_i with both focal elements of m_j will be A_i, since $A_j \cap A_i = \Theta \cap A_i = A_i$. Therefore these bpa's will not influence the calculation of $m_T(A_i)$, where $m_T(\cdot)$ is the result of the combination of all bpa's. When m_i is combined with m_k and A_k is below A_i in the tree the intersections will be A_k and A_i, respectively, and when A_k is neither above nor below A_i, the intersections will be empty and A_i, respectively. Thus, $m_T(A_i)$, $A_i \neq \Theta$, can be calculated as

$$m_T(A_i) = K \cdot m_i(A_i) \cdot \prod_{\forall j|\ A_i \not\subset A_j} m_j(\Theta)$$

where

$$K^{-1} = \sum_i m_i(A_i) \cdot \prod_{\forall j | A_i \not\subset A_j} m_j(\Theta).$$

Finally, the disconfirming bpa's are combined one by one with m_T. When belief is assigned to a subset, X, that is not in the tree this belief is reassigned to the smallest subset, A_i, such that X is a proper subset of A_i, $X \subset A_i$.

When a disconfirming bpa with focal element A_j^c is combined with m_T, we have the following.

- For subsets, X, equal to or below A_j in the tree, belief will only be assigned to X through its intersection with Θ, i.e. $Y \cap A_j^c \neq X$. No belief will be assigned to $X \cap A_j^c$ since this intersection is empty.

- For subsets, X, above A_j in the tree $X \cap A_j^c$ may or may not belong to the tree. If $X \cap A_j^c$ is in the tree then this subset will be assigned belief. If on the other hand $X \cap A_j^c$ is not in the tree then the belief assigned to $X \cap A_j^c$ is reassigned to the smallest subset, A_i, such that $X \cap A_j^c \subset A_i$. Since $X \cap A_j^c$ contains elements from both branches of X this smallest subset will always be X itself. Thus, X will be assigned belief from both intersections with the focal elements, i.e. $X \cap A_j^c = X \cap \Theta = X$.

- When the subset X is neither above, equal or below A_j, then the intersection between X and the focal element A_j^c is X itself. Besides the belief assigned by the intersection of X with the two focal elements of the disconfirming bpa, X will also be supported by the intersection of A_j^c with Y when $A_j \subset Y$ and $X \cup A_j = Y$.

Shafer and Logan [30] improved on the method by Gordon and Shortliffe. They showed that, while the algorithm by Gordon and Shortliffe usually produced a good approximation its performance was not as good when used with highly conflicting evidence. Besides not being approximate, the algorithm by Shafer and Logan also calculates belief for A_i^c of every partition, A_i, that is in the tree, thus it calculates the plausibility for all partitions in the tree. Both algorithms run in linear time. Interestingly, Shafer and Logan showed that the linear time complexity of their algorithm is linear in the number of the nonterminal nodes due to the local computations of their algorithm and linear in the tree's branching factor due to Barnett's approach.

The algorithm by Shafer and Logan can briefly be described as follows. First, for each subset in the tree, combine all the evidences for the subsets and against the subsets, respectively, Barnett style. The two resulting simple support functions for each node are then combined into a belief function with focal elements $\{A_i, A_i^c, \Theta\}$.

Then we propagate the belief in the tree.

- For each parent, A_i, of terminal nodes we combine the belief functions for all the children, store the belief in A_i and A_i^c at this parent node. This can be done with Barnett's method. These stored values will be used later when we propagate belief downwards through the tree. We combine the resulting belief function of the previous combination with the belief function for this parent and store the beliefs in A_i and A_i^c from the children and the parent at the parent node. These values will be used when we continue propagating belief upwards the tree. The procedure is repeated for the parents of these parents and so on climbing up the tree until we reach and store these values for the children of Θ.

- At the top we combine the belief functions from all the children of Θ and take one step down the tree to calculate the total belief in A_i and A_i^c for all children A_i of Θ.

- Finally, we climb down the tree step by step until we reach all terminal nodes and calculate on the way the total belief in A_i and A_i^c for all subsets A_i in the tree.

  First, in the last step down the tree we calculated the total belief in A_i and A_i^c for a particular parent. When climbing up the tree we stored with the same parent node the belief in A_i and A_i^c from the combination of all belief functions below the parent. From these values we can calculate the belief in A_i and A_i^c from all belief functions that are not below the parent.

  Secondly, for the parent, A_i, we once again combine all belief functions that are below the parent, as we did when we climbed up the tree, but this time we find the belief in A_j, A_j^c and $A_j \cup A_i^c$, where A_j is a child of A_i. This is again done with Barnett's method.

  Thirdly, the total belief in A_j and A_j^c for a child is now found by combining these two belief functions from the first and the second step, i.e. the belief function of all subsets below the parent with the belief function of all subsets not below the parent.

The algorithm by Shafer and Logan can handle evidence and calculate belief in partitions of the form $\{A_i, A_i^c\}$ for all subsets, A_i, in the tree. It can also calculate belief in partitions of the form $C_{A_i} \cup \{A_i^c\}$, where C_{A_i} is the set of children of A_i. However, their algorithm can not handle evidence for $C_{A_i} \cup \{A_i^c\}$. Since these two types of evidence correspond to data and domain knowledge respectively, this is a significant restriction. A generalization of the algorithm by Shafer and Logan that manages to take domain knowledge into account is the method for belief propagation in qualitative Markov trees by Shafer, Shenoy and Mellouli [31]. In a qualitative Markov tree the children are qualitatively conditionally independent (Mellouli *et al.* [21]) given the parent, i.e. in determining which element of a child is true, there is no additional information in knowing which element of another child is true once we know which element of the parent is true. Qualitative Markov trees can arise through constructing what Shafer, Shenoy and Mellouli call the tree of families and dichotomies. This is simply done by substituting each nonterminal node with subset A_i in a hierarchical tree by a parent-child pair with the dichotomy $\{A_i, A_i^c\}$ as subset at the parent and the family $C_{A_i} \cup \{A_i^c\}$ as subset at the child and furthermore substituting terminal nodes with subset A_i with the dichotomy $\{A_i, A_i^c\}$.

The algorithm for computing the total belief for every node in the tree of families and dichotomies is very simple.

- We will propagate belief to every neighbor, i.e. parent or child, A_j of every node A_i in the tree. We begin with the terminal nodes. Project the belief functions stored at every terminal node to its parent. For all neighbors but A_j, if belief functions have been projected to A_i from these neighbors then combine these belief functions and the belief function stored at A_i with Dempster's rule and project the result towards A_j. Whether or not belief has been projected from A_j to A_i is without significance to this rule.

- Finally, for every node A_i in the tree, when belief functions have been projected from all neighbors of A_j we calculate the total belief in A_i by combining all these projected belief functions and the belief function stored at A_i.

In [32] Shenoy and Shafer list the axioms under which local computations at the nodes are possible.

Shafer, Shenoy and Mellouli point out that this computational scheme reduces the time complexity from being exponential in the size of the frame to being exponential in the size of the largest partition.

1.3.4. Previous work on other topics

In any system it is of great importance to be able to construct explanations for the result. Such a methodology has been developed by Strat and Lowrance [35]. By using the discount operator we can for every body of evidence numerically calculate the derivative of belief and plausibility of any hypothesis. The more positive the derivative of support is, the more the body of evidence argues for the hypothesis and the more negative the derivative of plausibility is, the more the body of evidence argues against the hypothesis. Furthermore, it may, for example, be important to identify those evidences that strongly disagree with conclusions of an analysis. We can define consonance as

$$\mathrm{Cons}(m_\Theta) = \frac{1}{1 + \mathrm{Ent}(m_\Theta)}$$

where $\mathrm{Ent}(m_\Theta)$ is Yager's entropy [39];

$$\mathrm{Ent}(m_\Theta) = -\sum_{A_i \subseteq \Theta} m_\Theta(A_i) \cdot \log \mathrm{Pls}(A_i).$$

The more negative the derivative of consonance for an evidence is, the more that evidence disagrees with the overall results of the analysis.

The control problem discussed in Section 1.1 may be handled with evidential reasoning methods. For instance, Wesley [37] suggests that control knowledge sources may offer evidence about which control feature values have been observed. These control knowledge sources are similar to any other knowledge sources, except that their evidence concerns observations regarding the control of possible actions. The evidences may consist of a disjunction of several control feature values, each control feature value corresponding to a subset of all possible actions. After combining all evidence from the control knowledge sources with Dempster's rule we may extend the result from feature values to actions. Choosing the preferred action is now done by finding the action with the best evidential interval.

Finally, regarding other theories, both bayesian and possibility theories can be seen as subtheories of Dempster-Shafer theory (Klir [15]). When all focal elements are singletons the evidential interval of every proposition will have zero length, giving us a probability measure. This is when Dempster-Shafer theory reduces to Bayesian theory. On the other hand if all focal elements are ordered by set inclusion, i.e. $A_1 \subset A_2 \subset \ldots \subset A_n$, Dempster-Shafer theory reduces to Possibility theory with belief and plausibility as necessity and possibility measures, respectively.

2. OBJECTIVE OF THE WORK

The objective of this work is to develop methods that will offer decision support to anti-submarine warfare intelligence analysts (Bergsten *et al.* [3, 4]) through situation assessment based on a quantitative analysis of available intelligence reports, geographical knowledge and negative information from knowledge of intelligence forces whereabouts. More specifically, the objective is to support the intelligence analyst in those decisions that are based on knowledge, like number of units and previous, current and predicted positions of those units. Since every intelligence report contains information such as time, position, velocity, direction and submarine type, that might be uncertain and in fact every intelligence report itself comes with a general probability classification as to whether the report is true or false, these methods must be able to reason under uncertainty. Any such method should observe four guidelines of managing uncertainty in intelligence analysis (Gulick and Martin [10]):

- In intelligence analysis we are dealing with uncertain data which makes it fruitless to pursuit certainty in decision making.

- The source of uncertainty in data may arise directly from sensors or indirectly from inference of several sources or from interpretations. Whatever the case, this should be stated explicitly.

- We should apply an appropriate methodology to handle the uncertainty explicitly.

- Finally, the uncertainty of different interpretations must be communicated to the intelligence analyst. A suppression of uncertainty would result in a false sense of security.

It is obvious that this makes management of uncertainty one of the most important components of intelligence analysis.

3. SUMMARY OF ARTICLES I - V

This thesis contains five results regarding multiple target tracks in Evidential Intelligence Analysis. The first article [“On Nonspecific Evidence,” *Int. J. Intell. Syst.* **8**(6), 711-725, 1993] concerns the situation when we are reasoning with multiple events which should be handled independently. In this situation we may have evidences that are not only uncertain but their propositions may also be weakly specified in the sense that it may not be certain to which of many different events a proposition is referring. Thus, when a proposition is weakly specified it may be difficult to directly judge if it and a second proposition are referring to the same or to different events. If the proposition is not carrying any such information, this would be impossible.

Instead we will use the conflict between two propositions as an indication whether or not they are referring to the same event. This is an obvious choice since the conflict measures the lack of compatibility among evidences and propositions are more likely compatible when they are referring to the same event as compared to the situation when they are referring to different events where the actions are also most likely different. Consequently, the conflict will serve as a distance measure between bodies of evidence.

The distance measure may be used when partitioning the evidences into disjoint subsets. We will view the conflict within each subset not only as a measure of the lack of compatibility among evidences within the subset but also as an evidence against the current partitioning of the set of evidences, i.e. an evidence against the partitioning. This is acceptable since a critique against a part of the partitioning, the lack of compatibility among evidences, is a critique against the entire partitioning.

We may also have uncertainties in our domain knowledge. That is, our knowledge of the current number of events may only be probabilistic. Accordingly, we will also have a domain dependent conflict from a probability distribution about the number of events, partially conflicting with the actual number of events. This conflict will also be seen as an evidence against the current partitioning of the set of evidences.

The article establishes a criterion function of overall conflict when reasoning with multiple events. With this criterion we may handle evidences whose proposition is weakly specified. We will use the minimizing of overall conflict as the method of partitioning the set of evidences into subsets representing the events. This method will also handle the situation when the number of events are uncertain.

The criterion function is derived from viewing the conflicts within each subset and the domain conflict as evidences against the partitioning. Since the combination of these evidences will yield a zero belief in the partitioning, the most probable partitioning of evidences into disjoint subsets will correspond to the maximum of the plausibility of possible partitionings. Thus, the criterion function of overall conflict will be the difference, one minus the plausibility of possible partitionings. We will call this the metaconflict function.

The method of finding the best partitioning is based on an iterative minimization of the metaconflict function. Two theorems will significantly reduce the number of subsets for which we must find an optimal partitioning of evidences. Firstly, if we first find the optimal partitioning for the number of subsets where the domain conflict is smallest, r subsets, we need never consider any solutions with fewer subsets, and if we then find the optimal partitioning for j subsets where $j > r$, then we need never consider any further solutions where the number of subsets are fewer than j, etc. Secondly, if we have an optimal partitioning for some number of subsets we need never consider any solutions for some other number of subsets where the domain part of the entire metaconflict function is greater than the metaconflict of our best partitioning so far.

An algorithm for minimizing the metaconflict function is proposed. The proposed algorithm is based on the one hand on the two theorems and on the other hand on an iterative optimization among partitionings of evidence into disjoint subsets. This approach is proposed in order to avoid the combinatorial problem in minimizing the metaconflict function. In each step of the optimization the consequence of transferring an evidence from one subset to another is investigated.

After this, each subset of intelligence reports would then be referring to a different target and the reasoning can take place with each target treated separately.

In the second article ["Specifying Nonspecific Evidence," Manuscript] we extend the results of the previous article. Here we will not only search for the most plausible subset for each piece of evidence as was done in the first article. In addition we will now also specify each piece of evidence by finding the plausibility for every subset that the evidence belongs to the subset. This is done by observing changes in cluster and domain conflicts when a piece of evidence is moved out from, or brought into a subset.

We directly interpret the conflict in a subset as an evidence that there is at least one piece of evidence that is placed in the subset but does not actually belong to the subset; the first evidence. If some piece of evidence is taken out from the subset the conflict decreases. This decrease in conflict is interpreted as if there exists an evidence indicating that this piece of evidence that we took out does not belong to the subset; the second evidence. The remaining conflict in the subsets after this piece of evidence is taken out, is directly interpreted as another evidence indicating that there is at least one other piece of evidence that is placed in the remainder of the subset but does not actually belong to the remaining subset; the third evidence.

We will derive the second evidence that indicates that our piece of evidence does not belong to the subset. This is done by demanding that the belief in the proposition that "there is at least one piece of evidence that does not belong to the subset" should be equal in two different ways of looking at things. First, if that belief is based on the first evidence, before our piece of evidence is taken out from the subset, and secondly, if it is based on a combination of the second and third evidences, after it is taken out from the subset. Since we know the value of the first and third evidence and the second evidence is defined in such a way that a combination of the second and third evidence yields the first one, we may calculate the value of the second evidence.

Similarly, if our piece of evidence after it is taken out from the subset is brought into any another subset, its conflict will increase. This increase in conflict is interpreted as if there exists an evidence indicating that this piece of evidence does not belong to the new subset.

The domain conflict was interpreted as an evidence that there exists at least one piece of evidence that either does not belong to any of the subsets, or if this particular evidence happens to be placed in a subset by itself, as an evidence that it does belong to one of the other subsets. Thus, the domain conflict is an evidence indicating that the number of subsets is incorrect.

When a piece of evidence is in a subsets together with other evidences and it is taken out from the subset and put into a new subset by itself we have a change in the number of subsets and will observe an increase in domain conflict. We interpret this increase as if there exists an evidence indicating that our piece of evidence does not belong to this new subset.

Finally, if our piece of evidence is in a subset by itself and we take it out and move it to some other subset we also have a change in the number of subsets and we might have either an increase or a decrease in domain conflict. An increase is interpreted as if there exists an evidence that our piece of evidence does belong to the subset where it is placed and a decrease that it does not.

We will find such evidence regarding each piece of evidence and for every subset.

When this is done we make a partial specification of each piece of evidence. We combine all evidence from different subsets regarding this piece of evidence and calculate for each subset the belief and plausibility that our piece of evidence belongs to the subset. The belief in this will always be zero, with this exception, since every proposition states that our evidence does not belong to some subset. The exception is when our evidence is in a subset by itself and we receive an increase in domain conflict when it is moved to an other subset. That was interpreted as if there exists an evidence that our piece of evidence does belong to the subset where it is placed. Then we will also have a nonzero belief in that our piece of evidence belongs to the subset.

In the combination of all evidences regarding our piece of evidence we may receive support for a proposition stating that it does not belong to any of the subsets and can not be put into a subset by itself. That proposition is false and its support is the conflict in Dempster's rule, and also an indication that the evidence might be false.

In a subsequent reasoning process we will discount evidences based on their degree of falsity. If we had no indication as to the possible falsity of the evidence we would take no action, but if there existed such an indication we would pay ever less regard to the evidence the higher the degree was that the evidence is false and pay no attention to the evidence when it is certainly false. This is done by discounting the evidence with one minus the support of the false proposition.

Also, it is apparent that some evidences, due to a partial specification of affiliation, might belong to one of several different subsets. Such a piece of evidence is not so useful and should not be allowed to strongly influence the subsequent reasoning process within a subset.

If we plan to use an evidence in the reasoning process of some subset, we must find a credibility that it belongs to the subset in question. An evidence that cannot possible belong to a subset has a credibility of zero and should be discounted entirely for that subset, while an evidence which cannot possibly belong to any other subset and is without any support whatsoever against this subset has a credibility of one and should not be discounted at all when used in the reasoning process for this subset. That is, the degree to which an evidence can belong to a subset and no other subset corresponds to the importance the evidence should be allowed to play in that subset.

Here we should note that each original piece of evidence regardless of in which subset it was placed can be used in the reasoning process of any subset that it belongs to with a plausibility above zero, given only that it is discounted to its credibility in belonging to the subset.

When we begin our subsequent reasoning process in each subset, it will naturally be of vital importance to know to which event the subset is referring. This information is obtainable when the evidences in the subset have been combined. After the combination, each focal element of the final bpa will in addition to supporting some proposition regarding an action also be referring to one or more events where the proposed action may have taken place. Instead of summing up support for each event and every subset separately, we bring the problem to the metalevel where we simultaneously reason about all subsets, i.e. which subsets are referring to which events. In this analysis we use our domain knowledge stating that no more than one subset may be referring to an event. From each subset we then have an evidence indicating which events it might be referring to. We combining all the evidence from all different subsets with the restriction that any intersection in the combination that assigns one event to two different subsets is false. This method has a much higher chance to give a clearly preferable answer regarding which events is represented by which subsets, than that of only viewing the evidences within a subset when trying to determine its event.

The extension in this article of the methodology to partition nonspecific evidence developed in the first article imply that an evidence will now be handled similarly by the subsequent reasoning process in different subsets if these are of approximately equal plausibility for the evidence. Without this extension the most plausible subset would take the evidence as certainly belonging to the subset while the other subsets would never consider the evidence at all in their reasoning processes.

In the third article ["Finding a Posterior Domain Probability Distribution by Specifying Nonspecific Evidence," Manuscript] we extend the work of the previous two articles. We aim to find a posterior probability distribution regarding the number of subsets by combining a given prior distribution with evidence regarding the number of subsets that we received from the evidence specifying process.

We use the idea that each single piece of evidence in a subset supports the existence of that subset to the degree that this evidence supports anything at all other than the entire frame. In the evidence specifying process of the previous article we discounted each single evidence for its degree of falsity and its degree of credibility in belonging to the subset where it was placed. For each subset separately, we now combine all evidence within a subset and the resulting evidence is the total support for that subset.

The degree to which the resulting evidence from this combination in its turn supports anything at all other than the entire frame, is then the degree to which all the evidence within the subset taken together supports the existence of this subset, i.e. that it is a nonempty subset that belongs to set of all subsets.

For every original piece of evidence we derived in the previous article an evidence with support for a proposition stating that this piece of evidence does not belong to the subset. If we have such support for every single piece of evidence in some subset, then this is also support that the subset is false. In that case none of the evidences that could belong to the subset actually did so and the subset was derived by mistake. Thus, we will discount the just derived evidence that support the existence of the subsets for this possibility.

Such discounted evidences that support the existence of different subsets, one from each subset, are then combined. The resulting bpa will have focal elements that are conjunctions of terms. Each term give support in that some particular subset belongs the set of all subsets, i.e. that it is a nonempty subset.

From this we can create a new bpa that is concerned with the question of how many subsets we have. This is done by exchanging each and every proposition in the previous bpa that is a conjunction of r terms for one proposition in the new bpa that is on the form $|\chi| \geq r$, where χ is the set of all subsets. The sum of support of all focal elements in the previous bpa that are conjunctions of length r is then awarded the focal element in the new bpa which supports the proposition that $|\chi| \geq r$.

A proposition in the new bpa is then a statement about the existence of a minimal number of subsets. Thus, where the previous bpa is concerned with the question of which subsets have support, the new bpa is concerned with the question of how many subsets are supported. This new bpa gives us some opinion that is based only on the evidence specifying process, about the probability of different numbers of subsets.

In order to obtain the sought-after posterior domain probability distribution we combine this newly created bpa that is concerned with the number of subsets with our prior domain probability distribution which was given to us in the problem specification.

Thus, by viewing each evidence in a subset as support for the existence of that subset we were able to derive a bpa, concerned with the question of how many subsets we have, which we could combine with our prior domain probability distribution in order to obtain the sought-after posterior domain probability distribution.

The fourth article ["Dempster's Rule for Evidence Ordered in a Complete Directed Acyclic Graph," *Int. J. Approx. Reasoning* **9**(1), 37-73, 1993] derives a special case algorithm making it computationally feasible to analyze the possible tracks of a target. When it is uncertain whether or not the propositions of any two evidences are in logical conflict we may model the uncertainty by an additional evidence against the simultaneous belief in both propositions and treat the two original propositions as non-conflicting. This will give rise to a complete directed acyclic graph with the original evidences on the vertices and the additional ones on the edges, where the vertices represent states and the edges transitions between states. We may think of the vertices as positions in time and space, the edges as transitions between these positions and the sequence of states as a path along which some object may have moved. Transitions are only possible from a vertex with a lower index to one with a higher. The propositions of the evidences on the edges may, for example, tell us that the time difference between the states may be to small in relation to their distance.

We are interested in finding the most probable sequence of vertices through the graph, i.e. the most probable path. The frame of discernment is the set of all 2^n possible different paths through the graph of n vertices, where transitions are possible only from lower to higher ranked vertices. At every vertex we have evidence supporting the proposition that this vertex is included in the sequence. For every pair of vertices there is an edge between the vertices that is associated with an evidence expressing the degree of doubt about a direct transition between the two vertices. All the corresponding belief functions are simple support functions. Our interest is the problem where one begins with a basic probability assignment for all vertices and against all directed edges, i.e. evidence that the path does not include the first vertex of the edge or that it does not include the last vertex of the edge or that the path does include at least one vertex between the first and last vertices of the edge, thus excluding any direct transfer between the vertices.

The algorithm reasons about the logical conditions of a completely specified path through the complete directed acyclic graph. It is hereby gaining significantly in time and space complexity compared to the step by step application of Dempster's rule. Here, we are making the assumption that only one path at a time is permitted through the graph, i.e. two objects cannot pass through the graph at the same time. The problem of analyzing paths of multiple objects can be solved by the partitioning method of the first article after which we may reason about possible paths separately for each partitioning.

We will first explain the mathematical reasoning behind this algorithm, which calculates support and plausibility in the following steps: unnormalized plausibility, unnormalized support, conflict and finally support and plausibility normalized by the conflict.

Let us start with the unnormalized plausibility and see what is sufficient to make a path plausible. Plausibility for a path means to which degree a path is possible, i.e. to which degree no known factors speak against this path. There are only two types of items which speak against a path, the positive evidence for vertices that are not included in the path and the negative evidence against edges between vertices that are included in the path. This means that the degree to which we do not assign support to these evidences equals the degree to which the path is possible.

The algorithm for support is much more complicated than the one for plausibility. It is not only necessary to find out which evidence speaks against the path, it is also necessary to insist that the evidence of the vertices and edges that are included in the path speaks in favor of it.

We have the following conditions. Transitions between consecutive vertices in the path must be possible, i.e. the evidences that offer support against those edges disconfirms the path. Furthermore, no vertices outside the path are permitted to be visited. Thus, evidences that offer support for these vertices will also disconfirm the path. These are the same arguments as for the plausibility.

But even to the degree that these evidences do not disconfirm the path, we can not be sure that a vertex outside of the path is not visited. However, certain evidences will confirm the path. For instance, no transitions should be possible from vertices before the first vertex of the path to this vertex, i.e. all evidences against edges from vertices before the first vertex of the path to this vertex will confirm the path. Also, no transition should be possible from the last vertex of the path to vertices after this vertex. Furthermore, for the vertices not belonging to the path which are located between vertices of the path we state that either, no transition should be possible to these vertices from vertices in the path, or if such a transition is possible then it should not be possible to rejoin the path at its following vertex.

Finally, the vertices that belongs to the path offer confirmation to the path. The first and the last vertex of the path confirm the path by their proposition. For every intermediate vertex in the path there are two different possibilities. The evidence at the vertex confirms the path to some degree. To the degree the evidence does not offer support to the path, the path may still be supported if the evidence against edges is speaking against all other ways from the last vertex visited before this vertex to the first vertex visited after this vertex.

The time complexity of support and plausibility for a single completely specified path when measured in the size of the frame is $O(|\Theta| \cdot \log|\Theta|)$ compared to $O(|\Theta|^{\log|\Theta|})$ of the brute force application of Dempster's rule.

The fifth and final article ["On Rho in a Decision-Theoretic Apparatus of Dempster-Shafer Theory," Manuscript] concerns a decision-theoretic apparatus for Dempster-Shafer theory developed by Thomas M. Strat. In this apparatus, expected utility intervals are constructed for different choices. To find the preferred choice when the expected utility interval of one choice is included in that of another, it is necessary to interpolate a discerning point in the intervals. This is done by a parameter ρ. As a result, we will end up with a set of expected utility intervals ordered by interval inclusion, $[E_{1*},E_1^*] \subseteq [E_{2*},E_2^*] \subseteq \ldots \subseteq [E_{n*},E_n^*] \subseteq [0,1]$. Here we have renumbered the choices by the order of interval inclusion. When we have several choices they may be preferred in different intervals of ρ. Let us call the point where choices *i* and *j* are equally preferable ρ_{ij}. We already know that choice 1 is preferred when $\rho = 0$, since this choice has the highest lower expected utility among all choices, and it will remain the preferred choice while ρ is less than $\min_i \rho_{1i}$, the smallest of all ρ_{ij}'s and the first point of preference change. Beyond this point, choice *i* will be preferable over choice 1.

Continuing, choice *i* will now be preferred up to the point where $\rho = \min_j \rho_{ij}$, etc. Thus, the probability for the choices to be preferred are not equal. This probability varies with the length of the interval under which it is preferred.

In Strats apparatus, an assumption has to be made about the value of ρ. In this article we demonstrate that it is sufficient to assume a uniform probability distribution for ρ to be able to discern the most preferable choice. All we know about the value of ρ is that it is a parameter the belongs to the set of real numbers between 0 and 1, $\rho \in [0, 1]$, i.e. we know that our frame is that same set of numbers, $\Theta = [0, 1]$. However, we demand that the nonspecificity of our new distribution should be equal to the nonspecificity of the original assignment for any size of the frame. This means that we have only one continuous part of the probability distribution for ρ, and that it covers the entire interval from 0 to 1, i.e. a uniform probability distribution.

If we are only interested in a simple maximizing of utility then adopting a uniform probability distribution for ρ yields the same result as setting $\rho = 0.5$. Then, for simplicity, we might as well set $\rho = 0.5$ and choose the alternative that yields the highest expected utility as our decision.

In a situation with several different decision makers we might sometimes be more interested in having the highest expected utility among the decision makers rather than only trying to maximize our own expected utility. If the number of alternatives is equal to the number of decision makers then all we have to do is to choose the alternative that is preferred under the maximal interval length. That will be the choice with the highest probability of giving us the highest expected utility. When the number of decision makers are less than the number of choices the situation becomes much more complex. We must here take into account not only the choices already done be other decision makers but also the rational choices we can assume to be made by later decision makers. The preference of each alternative to some decision makers is shown to be the probability that the alternative has the highest expected utility after all decision makers have made their choices.

4. CONCLUSIONS

In supporting intelligence analysts in anti-submarine warfare we have developed a method based on Dempster-Shafer theory for analyzing tracks of multiple submarines. This method may be used to obtain general domain knowledge about operations as well as for analyzing a specific operation before making predictions.

The method partitions intelligence reports into subsets. Each subset is here representing a possible submarine. The method is based on minimizing a criterion function of overall conflict in partitions of intelligence reports. It is able to reason simultaneously about the optimal number of submarines, which may be uncertain, and the optimal partition of intelligence reports among the submarines. We made a partial specification of each intelligence report when it was uncertain to which submarine the report might belong. This specification was made by observing changes in cluster and domain conflicts when the intelligence report was moved from one subset to another. We also found probabilities regarding the number of submarines by combining our prior knowledge with results received from the specifying process.

Then, for each possible submarine a new algorithm calculates support and plausibility for possible tracks through the graph of positions of intelligence reports within the corresponding subset. When calculating the probability of a possible track we take account of both positive evidence of intelligence reports and negative evidence against the possibility of transitions between the positions of those reports. Finally, for each possible submarine its possible tracks are ranked according to their support, giving intelligence analysts a chance to reflect upon the current situation and its recent history before making predictions about the future.

By applying Dempster-Shafer theory to the field of intelligence analysis and in doing so giving the intelligence analysts the opportunity to directly investigate the partial specification of intelligence reports and to investigate and influence the partitioning of intelligence reports as well as all further results based on that partitioning we believe we have followed the four guidelines of managing uncertainty in intelligence analysis (Gulick and Martin [10]).

BIBLIOGRAPHY

1. Andersen, S. K., Olesen, K. G., Jensen, F. V., and Jensen, F., HUGIN - a Shell for Building Bayesian Belief Universes for Expert Systems, *Proceedings of the 11th International Joint Conference on AI*, Detroit, Mich., 1080-1085.
2. Barnett, J. A., Computational Methods for a Mathematical Theory of Evidence, *Proceedings of the 7th International Joint Conference on AI,* Vancouver, BC, 868-875, 1981.
3. Bergsten, U., Schubert, J., and Svensson, P., Beslutsstödssystemet Dezzy - En Översikt, in *Dokumentation 7 juni av Seminarium och fackutställning om samband, sensorer och datorer för ledningssystem till försvaret, MILINF 89,* Försvarets materielverk, Enköping, July 5-9, 1989, Telub AB, Växjö, 1989, 07B2:19-07B2:31. FOA report B 20078-2.7, Department of Weapon Systems, Effects and Protection, National Defence Research Establishment, Stockholm, 1990.
4. Bergsten, U., Schubert, J., and Svensson, P., Dezzy - ett Demonstrationssystem för Analys av Ubåtsunderrättelser, in Delprojekt Informationssystem inom Huvudprojekt Ubåtsskydd - Slutrapport (Schubert, J., Ed.), FOA report A 20046-2.7, Department of Weapon Systems, Effects and Protection, National Defence Research Establishment, Stockholm, 8-10, 1990.
5. Chiu, D. K. Y., and Wong, A. K. C., Synthesizing Knowledge: A Cluster Analysis Approach Using Event Covering, *IEEE Trans. Syst. Man, Cybern.* **16**(2), 251-259, 1986.
6. Cohen, P. R., The Control of Reasoning Under Uncertainty: A Discussion of Some Programs, *The Knowledge Eng. Rev.* **2**(1), 1987.
7. Cohen, P. R., Architectures and Strategies for Reasoning Under Uncertainty, in *Readings in Uncertain Reasoning,* The Morgan Kaufmann Ser. in Representation and Reasoning, (G. Shafer and J. Pearl, Eds.), Morgan Kaufmann Publ., San Mateo, Calif., 167-176, 1990.
8. Dempster, A. P., A generalization of Bayesian inference, *J. Roy. Stat. Soc. Ser. B* **30**(2), 205-247, 1968.
9. Gordon, J., and Shortliffe, E. H., A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space, *Artif. Intell.* **26**(3), 323-357, 1985.
10. Gulick, R. M., and Martin, A. W., Managing Uncertainty in Intelligence Data - An Intelligence Imperative, in *Science of Command and Control: Coping With Uncertainty* (S. E. Johnson and A. H. Levis, Eds), AFCEA International Press, Washington D.C., 10-18, 1988.
11. Jain, A. K., and Dubes, R. C., *Algorithms for Clustering Data,* Prentice-Hall, Englewood Cliffs, N.J., 1988.
12. Kennes, R., and Smets, P., Computational Aspects of the Möbius Transformation, in *Uncertainty in Artificial Intelligence 6,* Machine Intelligence and Pattern Recognition Ser., Vol. 12(P.P. Bonissone, M. Henrion, L.N. Kanal and J.F. Lemmer, Eds.), Elsevier Science Publ. (North-Holland), Amsterdam, 401-416, 1991.

13. de Kleer, J., An Assumption-Based Truth Maintenance System, *Artif. Intell.* **28**, 127-162, 1986.
14. Kleyle, R., and de Korvin, A., A Belief Function Approach to Information Utilization in Decision Making, *J. Am. Soc. Inf. Sci.* **41**(8), 569-580, 1990.
15. Klir, G. J., Probabilistic Versus Possibilistic Conceptualization of Uncertainty, in *Analysis and Management of Uncertainty: Theory and Applications,* Machine Intelligence and Pattern Recognition Ser., Vol. 13(B.M. Ayyub, M.M. Gupta and L.N. Kanal, Eds.), Elsevier Science Publ. (North-Holland), Amsterdam, 13-25, 1992.
16. Laskey, K. B., and Lehner, P. E., Belief Maintenance: An Integrated Approach to Uncertainty Management, *Proceedings of the 7th National Conference on AI,* Saint Paul, Minn., 210-214, 1988.
17. Laskey, K. B., and Lehner, P. E., Assumptions, Beliefs and Probabilities, *Artif. Intell.* **41**(1), 65-77, 1989.
18. Lesh, S. A., *An Evidential Theory Approach to Judgment-Based Decision Making,* Ph.D. Thesis, Dept. of Forestry and Environmental Studies, Duke University, Durham, N.C., 1986.
19. Lowrance, J. D., and Garvey, T. D., Evidential Reasoning: An Implementation for Multisensor Integration, Technical Note 307, Artificial Intelligence Center, Computer Science and Technology Division, SRI International, Menlo Park, Calif., 1983.
20. Lowrance, J. D., Garvey, T. D., and Strat, T. M., A framework for evidential-reasoning systems, *Proceedings of the 5th National Conference on AI,* Philadelphia, Penn., 896-901, 1986.
21. Mellouli, K., Shafer, G., and Shenoy, P. P., Qualitative Markov Networks, in *Uncertainty in Knowledge-Based Systems,* Lecture Notes in Computer Science, Vol. 286(B. Bouchon and R.R. Yager, Eds.), Springer-Verlag, Berlin, 69-74, 1987.
22. Nguyen, H. T., and Walker, E. A., On decision making using belief functions, in *Advances in the Dempster-Shafer Theory of Evidence* (R. Y. Yager, M. Fedrizzi and J. Kacprzyk, Eds.), John Wiley & Sons, Inc., New York, 311-330, 1994.
23. Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible inference,* Morgan Kaufmann Publ., San Mateo, Calif., 1988.
24. Pearl, J., Reasoning with Belief Functions: An Analysis of Compatibility, *Int. J. Approx. Reasoning* **4**(5/6), 363-389, 1990.
25. Rasoulian H., Thompson W., Kazda L. F., and Parra-Loera R., A multisensor data fusion model for multitarget tracking, P*roceedings of the SPIE Conference on Hybrid Image and Signal Processing II,* Vol. 1297, Orlando, Fla., 24-34, 1990.
26. Shafer, G., *A Mathematical Theory of Evidence,* Princeton University Press, Princeton, N.J., 1976.
27. Shafer, G., Constructive Decision Theory, Working Paper, School of Business, Univ. of Kansas, Lawrence, Kan., 1982.
28. Shafer, G., Perspectives on the Theory and Practice of Belief Functions, *Int. J. Approx. Reasoning* **4**(5/6), 323-362, 1990.

29. Shafer, G., Rejoinders to Comments on "Perspectives on the Theory and Practice of Belief Functions", *Int. J. Approx. Reasoning* **6**(3), 445-480, 1992.
30. Shafer, G., and Logan, R., Implementing Dempster's Rule for Hierarchical Evidence, *Artif. Intell.* **33**(3), 271-298, 1987.
31. Shafer, G., Shenoy, P. P., and Mellouli, K., Propagating Belief Functions in Qualitative Markov Trees, *Int. J. Approx. Reasoning* **1**(4), 349-400, 1987.
32. Shenoy, P. P., and Shafer, G., Axioms for Probability and Belief-Function Propagation, in *Uncertainty in Artificial Intelligence 4,* Machine Intelligence and Pattern Recognition Ser., Vol. 9(R.D. Shachter, T.S. Levitt, L.N. Kanal and J.F. Lemmer, Eds.), Elsevier Science Publ. (North-Holland), Amsterdam, 169-198, 1990.
33. Smets, P., and Kennes, R., The transferable belief model, *Artif. Intell.* **66**(2), 191-234, 1994.
34. Strat, T. M., Decision Analysis Using Belief Functions, *Int. J. Approx. Reasoning* **4**(5/6), 391-417, 1990.
35. Strat, T. M., and Lowrance, J. D., Explaining Evidential Analyses, *Int. J. Approx. Reasoning* **3**(4), 299-353, 1989.
36. Thoma, H. M., *Factorization of Belief functions,* Doctoral Dissertation, Dept. of Statistics, Harvard Univ., Cambridge, Mass., 1989.
37. Wesley, L., Reasoning About Control: The Investigation of an Evidential Approach, *Proceedings of the 8th International Conference on AI,* Karlsruhe, 203-206, 1983.
38. Xu, H., An Efficient Tool for Reasoning with Belief Functions, *Proceedings of the 4th International Conference on Information Processing and Management of Uncertainty,* Mallorca, 65-68, 1992.
39. Yager, R., Entropy and Specificity in a Mathematical Theory of Evidence, *Int. J. Gen. Syst.* **9**, 249-260, 1983.
40. Zarley, D., Hsia, Y.-T., and Shafer, G., Evidential Reasoning Using DELIEF, *Proceedings of the 7th National Conference on AI,* Saint Paul, Minn., 205-209, 1988.

On Nonspecific Evidence

Int. J. Intell. Syst. **8**(6), 711-725, 1993.

I

On Nonspecific Evidence

Johan Schubert
Division of Applied Mathematics and Scientific Data Processing, Department of Weapon Systems, Effects and Protection, National Defence Research Establishment, S-172 90 Sundbyberg, Sweden

When simultaneously reasoning with evidences about several different events it is necessary to separate the evidence according to event. These events should then be handled independently. However, when propositions of evidences are weakly specified in the sense that it may not be certain to which event they are referring, this may not be directly possible. In this article a criterion for partitioning evidences into subsets representing events is established. This criterion, derived from the conflict within each subset, involves minimizing a criterion function for the overall conflict of the partition. An algorithm based on characteristics of the criterion function and an iterative optimization among partitionings of evidences is proposed.

I. INTRODUCTION

A problem of major importance when reasoning with uncertainty is that in many situations evidences will not only be uncertain but their propositions may also be weakly specified in the sense that it may not be certain to which event a proposition is referring. In some cases the propositions may not carry any such information, making it impossible to differentiate between different events. Furthermore, the domain knowledge regarding events may be uncertain. For instance, our knowledge of the current number of events may only be probabilistic.

When reasoning about some proposition it is crucial not to combine evidences about different events in the mistaken belief that they are referring to the same event. For this reason every proposition's action part must be supplemented with an event part describing to which event the proposition is referring. The event part may be more or less weakly specified dependent on the evidence. If the evidences could be clustered into subsets representing events that should be handled separately from the others the situation would become manageable.

A simple example will illustrate the terminology. Let us consider the burglaries of two bakers' shops at One and Two Baker Street, event 1 (E_1) and event 2 (E_2), i.e., the number of events is known to be 2. One witness hands over an evidence, specific with respect to event, with the proposition: ''The

INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, VOL. 8, 711–725 (1993)

CCC 0884-8173/93/060711-15

burglar at One Baker Street,'' event part: E_1, ''was probably brown haired (B),'' action part: B. A second anonymous witness hands over a nonspecific evidence with the proposition: ''The burglar at Baker Street,'' event part: E_1, E_2, ''might have been red haired (R),'' action part: R. That is, for example:

evidence 1:	evidence 2:
proposition:	proposition:
action part: B	action part: R
event part: E_1	event part: E_1, E_2
$m(B) = 0.8$	$m(R) = 0.4$
$m(\Theta) = 0.2$	$m(\Theta) = 0.6$

The aim of this article is to establish, within the framework of Dempster–Shafer theory,[1–3] a criterion function[4] of overall conflict when reasoning with multiple events. With this criterion we may handle evidences whose proposition is weakly specified in its event part. We will use the minimizing of overall conflict as the method of partitioning the set of evidences into subsets representing the events. This method will also handle the situation when the number of events are uncertain.

An algorithm for minimizing the overall conflict will be proposed. The proposed algorithm is based, on the one hand, on characteristics of the criterion function for varying number of subsets and, on the other hand, on an iterative optimization among partitionings of evidence for a fixed number of subsets.

This algorithm was developed as a part of a multiple-target tracking algorithm for an antisubmarine intelligence analysis system.[5,6] In this application a sparse flow of intelligence reports arrives at the analysis system. These reports carry a proposition about the occurrence of a submarine at a specified time and place, a probability of the truthfulness of the report, and may contain additional information such as velocity, direction, and type of submarine.

The intelligence reports are never labeled as to which submarine they are referring to but it is of course possible to differentiate between two different submarines of two intelligence reports if the reports are known to be referring to different types of submarines. Moreover, times and positions of two different reports may be such that it is impossible to travel between the two positions at their respective times and therefore possible to differentiate between the two submarines. However, when this is not the case differentiation will not be possible. Instead we will use the conflict between the two intelligence reports as a probability that the two reports are referring to different submarines.

Before analyzing the possible tracks for an unknown number of submarines we want to separate the intelligence reports into subsets according to which submarine they are referring to and then analyze the possible tracks for each submarine separately. The most probable partition of reports into subsets is done by minimizing the criterion function of overall conflict.

In this application the action part of the intelligence report proposition states that a submarine was at the indicated time and position while the event part of a report is informal, often weakly specified and contained in the information that to some degree separates reports, such as time and position of the other reports, nonfiring sensors placed between reports, etc.

In Sec. II we will discuss how to use the overall conflict for separating nonspecific evidences. Following this, the criterion function for overall conflict of multiple events will be investigated (Sec. III). We then discuss the behavior of the criterion function in iterative optimization (Sec. IV). Finally, we propose an algorithm for partitioning evidences into subsets (Sec. V), based on the criterion function and a hill-climbing–like iterative optimization. We conclude with a detailed example (Sec. VI).

II. SEPARATING NONSPECIFIC EVIDENCE

When we have evidences with conflicting event parts we would like to separate them into disjoint subsets. After this, the reasoning should take place with the evidences in each subset treated separately. However, when the event part of a proposition is weakly specified with respect to which of many different events it is referring, it may be difficult if not impossible to directly judge whether or not it and a second proposition are referring to the same event. If, for instance, the first proposition is referring to events 1 or 2 and the second proposition is referring to events 1 or 3 it is uncertain whether or not they are referring to the same event. Thus, it will not be possible to separate evidences based only on their proposition's event parts.

Instead we will separate evidences by their conflict. This is an obvious choice since the conflict measures the lack of compatibility among evidences and the action parts of propositions are more likely compatible when they are referring to the same event as compared to the situation when they are referring to different events where the actions are also most likely different. Evidences are considered conflicting when they have empty intersections between representations of the proposition action parts with identical specific proposition event parts, i.e., propositions certainly referring to one and the same event. However, since all calculations take place within subsets where the evidences are presumed to be referring to the same event, we will have a conflict in two different situations. Firstly, we have a conflict if the proposition action parts are conflicting regardless of the proposition event parts since they are presumed to be referring to the same event. Secondly, if the proposition event parts are conflicting then, regardless of the proposition action parts, we have a conflict with the presumption that they are referring to the same event. In order to avoid that evidences with specific and identical event parts end up in different subsets we may precombine these evidences and henceforth handle them as one evidence. The idea of using the conflict as distance measure between bodies of evidence has been suggested earlier by Lowrance and Garvey[7] and by Lesh.[8]

The conflict within each subset will not only be seen as a measure of the lack of compatibility among evidences within the subset but also as an evidence against the current partitioning of the set of evidences, χ, into the subsets χ_i. This is an intuitively correct definition since a critique against a part of the partitioning, the lack of compatibility among evidences, is a critique against the entire partitioning, i.e., an evidence against the partitioning. A zero conflict is no evidence against the partitioning and a conflict of one is an evidence that the partitioning is impossible. The frame of discernment is here $\Theta = \{$ Partition, $\neg$Partition$\}$. That is, the basic probability assignment against the partitioning from subset χ_i is

$$m_{\chi_i}(\neg\text{Partition}) \triangleq \text{Conf}(\{e_j | e_j \in \chi_i\}),$$

$$m_{\chi_i}(\Theta) \triangleq 1 - \text{Conf}(\{e_j | e_j \in \chi_i\})$$

where $\{e_j | e_j \in \chi_i\}$ is the set of evidences belonging to subset χ_i and Conf$(\cdot)$ is the conflict, k, in Dempster's rule.

When the evidences are not simple support functions the conflict measure might, at first glance, seem odd as a distance measure between bodies of evidence, since two nonsimple support functions with identical sets of focal elements may have a nonzero conflict. However, this need not be nonintuitive, as shown by the case of four simple support functions, the two first identical and in conflict with the two identical remaining simple support functions. If the two first and the two last functions are combined then a conflict measure with the intuitive properties of a distance measure is obtained, when the two resulting simple support functions are combined. If, on the other hand, we combine the first with the third and the second with the fourth, we receive two identical nonsimple support functions whose combination will result in a nonzero conflict. Clearly, if the conflict measure was intuitive as distance measure in the first combination order then it is also intuitive in the second. Then, at least for support functions that are derivable from simple support functions, it is not nonintuitive to have a nonzero distance measure for support functions with identical sets of focal elements.

In addition there will also be a domain-dependent conflict from a probability distribution about the number of subsets, E, conflicting with the actual current number of subsets, $\#\chi_i$. This conflict will also be seen as an evidence against the current partitioning of the set of evidences into the subsets,

$$m_D(\neg\text{Partition}) \triangleq \text{Conf}(\{E_i\}, \#\chi_i),$$

$$m_D(\Theta) \triangleq 1 - \text{Conf}(\{E_i\}, \#\chi_i).$$

Fusing these evidences with Dempster's rule yields

$$m(\neg\text{Partition}) = 1 - [1 - m_D(\neg\text{Partition})] \cdot \prod_{i=1}^{r} [1 - m_{\chi_i}(\neg\text{Partition})],$$

$$m(\Theta) = [1 - m_D(\neg\text{Partition})] \cdot \prod_{i=1}^{r} [1 - m_{\chi_i}(\neg\text{Partition})]$$

with belief and plausibility of the partitioning being

$$\text{Bel(Partition)} = 0,$$

$$\text{Pls(Partition)} = [1 - m_D(\neg\text{Partition})] \cdot \prod_{i=1}^{r} [1 - m_{\chi_i}(\neg\text{Partition})].$$

Finding the most probable partitioning of evidences into disjoint subsets representing different events will then be the problem of maximizing the plausibility of possible partitionings, or the dual problem of minimizing one minus the plausibility. The difference, one minus the plausibility of a partitioning, will be called the metaconflict of the partitioning.

III. METACONFLICT AS A CRITERION FUNCTION

Let us define the metaconflict function, derived in the preceding section, whose minimization per definition leads to the optimal partitioning of evidences into disjoint subsets.

DEFINITION. *Let the* metaconflict function,

$$Mcf(r, e_1, e_2, \ldots, e_n) \triangleq 1 - (1 - c_0) \cdot \prod_{i=1}^{r} (1 - c_i), \tag{1}$$

be the conflict against a partitioning of n evidences of the set χ *into r disjoint subsets* χ_i *where*

$$c_0 = \sum_{i \neq r} m(E_i) \tag{2}$$

is the conflict between r subsets and propositions about possible different number of subsets and

$$c_i = \sum_{I | \cap I = \varnothing} \prod_{e_j^k \in I} m(e_j^k)$$

is the conflict in subset i, where $I = \{e_j^k | e_j \in \chi_i\}$ *is a set of one focal element from the support function of each evidence in* χ_i.

Some characteristics of the metaconflict function will be useful when choosing the number of subsets of χ for which we must find an optimal partitioning of evidences.

The first theorem below states that if we have an optimal partitioning for

r subsets, then we need never consider any solutions with fewer than r subsets when the basic probability number for r subsets is greater than the basic probability number for fewer subsets. These solutions need never be considered because the nondomain part of the metaconflict function always increases with fewer subsets and when the basic probability number for fewer subsets is smaller than the basic probability number for r subsets, then the domain part of the metaconflict function for fewer subsets has also increased, yielding an overall increase in the metaconflict. The significance of this theorem is that it can be applied iteratively. If we first find the optimal partitioning for the number of subsets where $m(E_r)$ is greatest, we need never consider any solutions with fewer subsets than r, and if we then find the optimal partitioning for the greatest $m(E_j)$ where $j > r$, then we need never consider any further solutions where the number of subsets are fewer than j, etc.

The second theorem states that if we have an optimal partitioning for some number of subsets we need never consider any solutions for some other number of subsets where the domain part of the metaconflict function is greater than the metaconflict of our present partitioning. This theorem will also be used iteratively as we gradually find better optimizations, step by step eliminating some of the possible solutions where the number of subsets is greater than with our present partitioning.

Together, these two theorems will significantly reduce the number of iterative optimizations we must carry through for different numbers of subsets.

THEOREM 1. *For all j with $j < r$, if $m(E_j) < m(E_r)$ then $\min Mcf(r, e_1, e_2, \ldots, e_n) < \min Mcf(j, e_1, e_2, \ldots, e_n)$.*

Proof. From the fact $m(E_j) < m(E_r)$ and (2) it follows

$$c_0 = \sum_{i \neq r} m(E_i) = \sum_{i \neq j} m(E_i) + m(E_j) - m(E_r) < \sum_{i \neq j} m(E_i) = c_0'. \tag{3}$$

From (3) and by the definition of metaconflict (1) it is sufficient that

$$\forall j. \max \prod_{i=1}^{j} (1 - c_i') \leq \max \prod_{i=1}^{r} (1 - c_i)$$

for min Mcf($r, e_1, e_2, \ldots, e_n$) to be less than min Mcf($j, e_1, e_2, \ldots, e_n$). This is equivalent with

$$\forall j. \max \prod_{i=1}^{j} (1 - c_i') \leq \max \prod_{i=1}^{j+1} (1 - c_i).$$

It is sufficient to show that the partition into j subsets that yields the maximum is less than or equal to any partition into $j + 1$ subsets.

Let the partition into the j first of the $j + 1$ disjoint subsets be unchanged from the optimal partition into j subsets with the exception that one evidence is moved from one of the subsets with more than one evidence, say χ_k, to subset χ_{j+1}. There is, with only one evidence, no conflict in χ_{j+1}, $c_{j+1} = 0$.

Then

$$\prod_{i=1}^{j+1}(1 - c_i) = \prod_{i=1}^{j}(1 - c_i) = (1 - c_k) \cdot \prod_{\substack{i=1\\ \neq k}}^{j}(1 - c_i)$$
$$= (1 - c_k) \cdot \prod_{\substack{i=1\\ \neq k}}^{j}(1 - c_i') = \frac{1 - c_k}{1 - c_k'} \cdot \prod_{i=1}^{j}(1 - c_i') \geq \prod_{i=1}^{j}(1 - c_i')$$

since the conflict in χ_k after moving out an evidence, c_k, is always less than or equal to the conflict before moving, c_k'.

THEOREM 2. *For all j, if min* $Mcf(r, e_1, e_2, \ldots, e_n) < \sum_{i \neq j} m(E_i)$ *then min* $Mcf(r, e_1, e_2, \ldots, e_n) <$ *min* $Mcf(j, e_1, e_2, \ldots, e_n)$.

Proof. From the condition of the theorem and by (1) we have

$$\min Mcf(r, e_1, e_2, \ldots, e_n) < \sum_{i \neq j} m(E_i)$$
$$= \frac{\min Mcf(j, e_1, e_2, \ldots, e_n) + \prod_{i \neq j}(1 - c_i) - 1}{\prod_{i \neq j}(1 - c_i)}$$
$$= 1 - \frac{1 - \min Mcf(j, e_1, e_2, \ldots, e_n)}{\prod_{i \neq j}(1 - c_i)}$$
$$< \min Mcf(j, e_1, e_2, \ldots, e_n).$$

There are also two theorems regarding the stability of an optimal solution, i.e., that the partition of the optimal solution can not self-splinter into new subsets, Theorem 3, and that the partition is invariant with respect to evidence incompatible with the partition, Theorem 4.

Since the nondomain part of the metaconflict function decreases with the number of subsets it is only the domain conflict part of the metaconflict that prevents the number of subsets to be equal to the number of evidences. Thus whether or not a partitioning of evidences is stable depends on the relation between these conflicts.

THEOREM 3. *A partitioning is stable, i.e., the metaconflict increases if any evidence is removed from its subset to form a new subset, if the relative change in domain conflict is higher than all relative conflict changes of the subsets.*

Proof. If an evidence e_q is removed from χ_i and included into a new subset χ_{r+1} the metaconflict would change to

$$Mcf^* = 1 - (1 - c_0^*) \cdot (1 - c_i^*) \cdot \prod_{k \neq i}(1 - c_k)$$
$$= 1 - (1 - c_0) \cdot (1 - c_i) \cdot \prod_{k \neq i}(1 - c_k)$$

$$- [(1 - c_0^*) \cdot (1 - c_i^*) - (1 - c_0) \cdot (1 - c_i)] \cdot \prod_{k \neq i} (1 - c_k)$$
$$= Mcf + [(c_0^* - c_0) + (c_i^* - c_i) - (c_0^* \cdot c_i^* - c_0 \cdot c_i)] \cdot \prod_{k \neq i} (1 - c_k)$$

The partition is stable if ∀i. Mcf* > Mcf. That is, if

$$\forall i. \frac{(c_0^* - c_0)}{1 - c_0} > \frac{(c_0 \cdot c_i - c_0^* \cdot c_i^*) - (c_i^* - c_i)}{1 - c_0}$$
$$> c_i^* - c_i > \frac{(c_i^* - c_i)}{1 - c_i}.$$

Finally, a new evidence which is uncompatible with all subsets and introduced into its own subset will not change the partition.

THEOREM 4. *If P is a unique optimal partition of the set of evidences,* χ, *and* e_{n+1} *a new evidence which is highly conflicting with each subset* χ_i *is introduced into its own subset* χ_{r+1}. *Then the optimal partition of* $\chi \cup \{e_{n+1}\}$ *is* $P/\{e_{n+1}\}$.

Proof. The optimal partition of the new set of evidences is found by minimizing Mcf*($r + 1, e_1, e_2, \ldots, e_{n+1}$). However, since e_{n+1} is introduced into its own subset χ_{r+1}, a subset without conflict, this can be rewritten as a function of the minimization of the old metaconflict, Mcf($r, e_1, e_2, \ldots, e_n$);

$$\min Mcf^*(r + 1, e_1, e_2, \ldots, e_{n+1})$$
$$= \min 1 - (1 - c_0^*) \cdot \prod_{i=1}^{r+1} (1 - c_i)$$
$$= \min 1 - (1 - c_0^*) \cdot \prod_{i=1}^{r} (1 - c_i)$$
$$= \min 1 - \frac{1 - c_0^*}{1 - c_0} \cdot (1 - c_0) \cdot \prod_{i=1}^{r} (1 - c_i)$$
$$= \min 1 - \frac{1 - c_0^*}{1 - c_0} \cdot [1 - Mcf(r, e_1, e_2, \ldots, e_n)]$$
$$= 1 - \frac{1 - c_0^*}{1 - c_0} \cdot [1 - \min Mcf(r, e_1, e_2, \ldots, e_n)].$$

That is, finding the optimal partition of $\chi \cup \{e_{n+1}\}$ when e_{n+1} is introduced into its own subset is done by finding an optimal partition of χ. Since there is only one such partition, P, the optimal partition of $\chi \cup \{e_{n+1}\}$ is $P/\{e_{n+1}\}$.

IV. CONDITION FOR ITERATIVE OPTIMIZATION

For a fixed number of subsets a minimum of the metaconflict function can be found by an iterative optimization among partitionings of evidences into different subsets. This approach is proposed in order to avoid the combinatorial problem in minimizing the metaconflict function. In each step of the optimization the consequence of transferring an evidence from one subset to another is investigated. If an evidence e_q is transferred from χ_i to χ_j then the conflict in χ_j, c_j increases to

$$c_j^* = c_j + \sum_{\substack{A_k \in \chi_j | A_k \neq \emptyset \\ e_q^p \in e_q | A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m(A_k).$$

with new focal elements and basic probability assignments

$$m^*(A_k) = m(A_k) \cdot \sum_{e_q^p \in e_q | A_k = A_k \cap e_q^p} m(e_q^p)$$

and

$$m^*(A_k \cap e_q^p) = m(A_k) \cdot \sum_{e_q^p \in e_q | A_k \neq A_k \cap e_q^p} m(e_q^p)$$

where $\{A_k\}$ are the focal elements before the transfer of e_q and $\{A_k, e_q^p \cap A_k\}$ are the focal elements after the transfer. The conflict in χ_i, c_i decreases to

$$c_i^* = c_i - \sum_{\substack{A_k \in \chi_i | A_k \neq \emptyset \\ e_q^p \in e_q | A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m^*(A_k)$$

where

$$m^*(A_k) = m(A_k) \bigg/ \sum_{e_q^p \in e_q | A_k = A_k \cap e_q^p} m(e_q^p)$$

Here, $\{A_k\}$ are the focal elements after the transfer of e_q and $\{A_k, e_q^p \cap A_k\}$ are the focal elements before the transfer. That is, we find the basic probability assignment of the focal elements as if the evidence was not included in χ_i and calculate the additional conflict created by transferring the evidence to χ_i. This additional conflict is then deducted from the conflict, c_i, to calculate the conflict after having transferred the evidence from χ_i, c_i^*. Given this, the metaconflict is changed to

$$\begin{aligned} Mcf^* &= 1 - (1 - c_0) \cdot (1 - c_i^*) \cdot (1 - c_j^*) \cdot \prod_{k \neq i,j} (1 - c_k) \\ &= 1 - (1 - c_0) \cdot \prod_k (1 - c_k) \end{aligned}$$

$$+ (1 - c_0) \cdot \left(\prod_k (1 - c_k) - (1 - c_i^*) \cdot (1 - c_j^*) \cdot \prod_{k \neq i,j} (1 - c_k) \right)$$
$$= Mcf + (1 - c_0) \cdot [(1 - c_i) \cdot (1 - c_j)$$
$$- (1 - c_i^*) \cdot (1 - c_j^*)] \cdot \prod_{k \neq i,j} (1 - c_k).$$

The transfer of e_q from χ_i to χ_j is favorable if Mcf* < Mcf. From the last expression, this is the case if

$$(1 - c_i) \cdot (1 - c_j) < (1 - c_i^*) \cdot (1 - c_j^*).$$

Rewriting this as

$$\frac{1 - c_j^*}{1 - c_j} > \frac{1 - c_i}{1 - c_i^*}$$

we substitute c_i^* and c_j^* with their expressions

$$\frac{1 - c_j - \sum\limits_{\substack{A_k \in \chi_j \mid A_k \neq \emptyset \\ e_q^p \in e_q \mid A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m(A_k)}{1 - c_j} > \frac{1 - c_i}{1 - c_i + \sum\limits_{\substack{A_k \in \chi_i \mid A_k \neq \emptyset \\ e_q^p \in e_q \mid A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m^*(A_k)}$$

which yields

$$1 - \frac{\sum\limits_{\substack{A_k \in \chi_j \mid A_k \neq \emptyset \\ e_q^p \in e_q \mid A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m(A_k)}{1 - c_j} > \frac{1}{1 + \dfrac{\sum\limits_{\substack{A_k \in \chi_i \mid A_k \neq \emptyset \\ e_q^p \in e_q \mid A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m^*(A_k)}{1 - c_i}}.$$

Finally, we conclude that the transfer of e_q from χ_i to χ_j is favorable if

$$\frac{\sum\limits_{\substack{A_k \in \chi_j \mid A_k \neq \emptyset \\ e_q^p \in e_q \mid A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m(A_k)}{1 - c_j} < \frac{\dfrac{\sum\limits_{\substack{A_k \in \chi_i \mid A_k \neq \emptyset \\ e_q^p \in e_q \mid A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m^*(A_k)}{1 - c_i}}{1 + \dfrac{\sum\limits_{\substack{A_k \in \chi_i \mid A_k \neq \emptyset \\ e_q^p \in e_q \mid A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m^*(A_k)}{1 - c_i}}.$$

Let us call these quotients ρ_j^q and ρ_i^q respectively, i.e., it is favorable to transfer e_q from χ_i to χ_j if $\rho_j^q < \rho_i^q$. It is, of course, most favorable to transfer e_q to χ_k, $k \neq i$, if $\forall j.\ \rho_k^q \leq \rho_j^q$. It should be remembered that this analysis concerns the situation where only one evidence is transferred from one subset to another. It may not be favorable at all to simultaneously transfer two or more evidences which are deemed favorable for individual transfer. It can easily be shown that when several different evidences are favorable to transfer it will be most favorable to transfer the evidence e_q that maximizes $(1 - \rho_k^q)/(1 - \rho_i^q)$.

V. AN ALGORITHM FOR MINIMIZING METACONFLICT

The algorithm for finding the partitioning of evidences among subsets that minimizes the metaconflict is based on Theorems 1 and 2 of the metaconflict function for finding the optimal number of subsets and an iterative optimization among partitionings of evidences for a fixed number of subsets. The iterative part of the algorithm, step 4 in the algorithm below, guarantees, like all hill-climbing algorithms, local but not global optimum.

Algorithm. Let S be the set of natural numbers less or equal to the number of evidences and T the empty set.

1. *Calculate* $\forall r.\ \Sigma_{i \neq r}\, m(E_i)$, *the conflict against a partitioning of the evidences into r subsets.*
2. *Let* $r = j \mid min_{j \in S}\, \Sigma_{i \neq j}\, m(E_i)$.
3. $T = T + \{r\}$, $S = S - \{r, j \mid j < r, j \in S\}$.
4. *Calculate min* $Mcf(r, e_1, e_2, \ldots, e_n)$.
    - 4.1. *Make an initial partition equal to the final partition of the last calculation of Mcf into the fist t subsets with the exception of moving the r − t most highly conflicting evidences from these subsets, updating the conflicts after each movement, one into each of the new r − t subsets. If it is the first calculation make any partition with at least one evidence in each subset. Calculate Mcf of the current partition.*
    - 4.2. *Let* $t = r$. *If* $r = 1$ *go to 4.5.*
    - 4.3. *For* $q = 1$ *to n. Suppose that* e_q *is currently in* χ_i.
        - 4.3.1. *If* $|\chi_i| = 1$ *go to 4.3 else calculate for* $1 \leq j \leq r$,

$$
\rho_j^q = \begin{cases}
\dfrac{\displaystyle\sum_{\substack{A_k \in \chi_j \mid A_k \neq \emptyset \\ e_q^p \in e_q \mid A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m(A_k)}{1 - c_j}, & j \neq i \\[4ex]
\dfrac{\dfrac{\displaystyle\sum_{\substack{A_k \in \chi_i \mid A_k \neq \emptyset \\ e_q^p \in e_q \mid A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m^*(A_k)}{1 - c_i}}{1 + \dfrac{\displaystyle\sum_{\substack{A_k \in \chi_i \mid A_k \neq \emptyset \\ e_q^p \in e_q \mid A_k \cap e_q^p = \emptyset}} m(e_q^p) \cdot m^*(A_k)}{1 - c_i}}, & j = i
\end{cases}
$$

4.4. *Transfer e_q from χ_i to χ_k, $k \neq i$, if*

$$\forall w. \frac{1-\rho_k^q}{1-\rho_i^q} \geq \frac{1-\rho_k^w}{1-\rho_i^w}$$

where $\forall j.\ \rho_k^w \leq \rho_j^w$.

4.5. *Update Mcf, c_i and c_k.*

4.6. *If Mcf is unchanged in step 4.5 then go to 5 else go to 4.3.*

5. $S = S - \{\forall j \mid j > r \mid \sum_{i \neq j} m(e_i^\chi) > \min Mcf(r, e_1, e_2, \ldots, e_n)\}$.

6. *If* $S \neq \emptyset$ *then go to 2 else answer* $\min_{t \in T} Mcf(t, e_1, e_2, \ldots, e_n)$.

VI. AN EXAMPLE

Let us, as an illustration of the problem solved in this article, consider a simple example of two possible burglaries, with a couple of evidences with simple support functions and some of the evidence weakly specified in the sense that it is uncertain to which possible burglary their propositions are referring. Assume that a baker's shop at One Baker Street has been burglarized, event 1. Let there also be some indication that a baker's shop across the street, at Two Baker Street, might have been burglarized, although no burglary has been reported, event 2. An experienced investigator estimates that a burglary has taken place at Two Baker Street with a probability of 0.4. We have received the following evidences. A credible witness reports that "a brown-haired man who is not an employee at the baker's shop committed the burglary at One Baker Street," evidence 1. An anonymous witness, not being aware that there might be two burglaries, has reported "a brown-haired man who works at the baker's shop committed the burglary at Baker Street," evidence 2. Thirdly, a witness reports having seen "a suspicious-looking red-haired man in the baker's shop at Two Baker Street," evidence 3. Finally, we have a fourth witness, this witness, also anonymous and not being aware of the possibility of two burglaries, reporting that the burglar at the Baker Street baker's shop was a brown-haired man. That is, for example:

evidence 1:
  proposition:
    action part: BO
    event part: E_1
  $m(BO) = 0.8$
  $m(\Theta) = 0.2$

evidence 2:
  proposition:
    action part: BI
    event part: E_1, E_2
  $m(BI) = 0.7$
  $m(\Theta) = 0.3$

evidence 3:
  proposition:
    action part: R
    event part: E_2
  $m(R) = 0.6$
  $m(\Theta) = 0.4$

evidence 4:
  proposition:
    action part: B
    event part: E_1, E_2
  $m(B) = 0.5$
  $m(\Theta) = 0.5$

domain probability distribution:

$$m(E_i) = \begin{cases} 0.6, i = 1 \\ 0.4, i = 2 \,. \\ 0, i \neq 1, 2 \end{cases}$$

Let us use the algorithm in Sec. V to investigate whether we have a one- or two-event problem and possibly separate the set of evidences into two disjoint subsets.

Algorithm: Let $S = \{1, 2, 3, 4\}$, $T = \emptyset$, where S are possible numbers of subsets and T different numbers of subsets for which we have minimized Mcf.
Step 1: We calculate the domain conflict from the probability distribution,

$$c_0 = \begin{cases} 0.4, r = 1 \\ 0.6, r = 2. \\ 1, r \neq 1, 2 \end{cases}$$

Step 2: The domain conflict is minimal for one subset, $r = 1$.
Step 3: Update $T := T + \{1\} = \{1\}$, $S := S - \{1\} = \{2, 3, 4\}$.
Step 4:
Step 4.1: In the initial partition all evidences are brought into one subset $\chi_1 = \{e_1, e_2, e_3, e_4\}$. Mcf = 0.884.
Step 4.2: $t = 1$. Since we only have one subset, $r = 1$, no transfers are possible, we go to 4.5.
Step 4.5: No evidences has been transferred, Mcf and c_1 is unchanged.
Step 4.6: Since Mcf was unchanged in step 4.5 we go to 5.
Step 5: We update, $S := S - \{3, 4\} = \{2\}$.
Step 6: Since $S \neq \emptyset$ there might exist better solutions, we go to 2.
Step 2: We minimize Mcf for two subsets, $r = 2$.
Step 3: Update, $T := T + \{2\} = \{1, 2\}$, $S := S - \{2\} = \emptyset$.
Step 4:
Step 4.1: As the initial partitioning move the most highly conflicting evidence from χ_1 to χ_2. We have e_1: $\rho_1^1 = 0.604$, e_2: $\rho_1^2 = 0.578$, e_3: $\rho_1^3 = 0.559$, e_4: $\rho_1^4 = 0.085$, i.e., we move e_1 from subset χ_1 to subset χ_2. Mcf = 0.804.
Step 4.2: $t = 2$.
Step 4.3: Since e_1 is in χ_2 and $|\chi_2| = 1$, e_1 can not be moved out of χ_2 and no ρ_j^1's are calculated. For $q = \{2, 3, 4\}$ we get for, e_2: $\rho_1^2 = 0.3$, $\rho_2^2 = 0.56$, e_3: $\rho_1^3 = 0.51$, $\rho_2^3 = 0.48$, e_4: $\rho_1^4 = 0.155$, $\rho_2^4 = 0$.
Step 4.4: We get for, e_2: $(1 - \rho_1^2)/(1 - \rho_1^2) = 1$, e_3: $(1 - \rho_2^3)/(1 - \rho_1^3) = 1.061$, e_4: $(1 - \rho_2^4)/(1 - \rho_1^4) = 1.184$, i.e., we move e_4 from χ_1 to χ_2. We get $\chi_1 = \{e_2, e_3\}$ and $\chi_2 = \{e_1, e_4\}$.
Step 4.5: We update conflicts, Mcf = 0.768, $c_1 = 0.42$, $c_2 = 0$.
Step 4.6: Mcf has changed and we continue at 4.3.
Step 4.3: Since $|\chi_1| = |\chi_2| = 2$ all evidences can be moved. For $q = \{1, 2, 3, 4\}$ we get for e_1: $\rho_1^1 = 0.634$, $\rho_2^1 = 0$, e_2: $\rho_1^2 = 0.42$, $\rho_2^2 = 0.48$, e_3: $\rho_1^3 = 0.42$, $\rho_2^3 = 0.54$, e_4: $\rho_1^4 = 0.155$, $\rho_2^4 = 0$.

Step 4.4: For all q we have $\rho_k^q = \rho_l^q$, i.e., no evidences are transferred.
Step 4.5: Mcf, c_1 and c_2 are unchanged.
Step 4.6: Since Mcf is unchanged, go to 5.
Step 5: $S := \emptyset$, there are no further possible solutions.
Step 6: Since $S = \emptyset$ we answer $\{\chi_1, \chi_2, \ldots, \chi_t\}$ where $t \in T = \{1, 2\}$ minimizes the metaconflict function,

$$\min_{t \in T} Mcf(t, e_1, e_2, \ldots, e_n),$$

i.e., we answer $\{\chi_1, \chi_2\}$ where $\chi_1 = \{e_2, e_3\}$, $\chi_2 = \{e_1, e_4\}$ for $t = 2$. Concluding, we see from the event parts of the evidences in each subset that χ_1 corresponds to event 2 and χ_2 corresponds to event 1.

VII. CONCLUSIONS

A criterion function of overall conflict has been established within the framework of Dempster–Shafer theory. An algorithm has been proposed for partitioning nonspecific evidence into subsets, each subset representing a separate event. The algorithm has a theoretical foundation in the minimizing of overall conflict of the partition when viewing the conflict within each subset as an evidence against the partition. The algorithm will not only be able to reason about the optimal partition of nonspecific evidence for a fixed number of events, it will also be able to reason simultaneously about the optimal number of events, which may be uncertain. An obvious drawback is the algorithm's inability to guarantee global optimality.

I would like to thank Stefan Arnborg, Ulla Bergsten, and Per Svensson for their helpful comments regarding this article.

References

1. A.P. Dempster, "A generalization of Bayesian inference," *J. R. Stat. Soc. Ser. B*, **30,** 205–247 (1968).
2. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University, Princeton, 1976).
3. G. Shafer, "Perspectives on the Theory and Practice of Belief Functions," *Int. J. Approx. Reason.*, **4,** 323–362 (1990).
4. A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data* (Prentice-Hall, Englewood Cliffs, NJ, 1988).
5. U. Bergsten, J. Schubert, and P. Svensson, "Beslutsstödssystemet Dezzy—En Översikt," In *Dokumentation 7 juni av Seminarium och fackutställning om samband, sensorer och datorer för ledningssystem till försvaret, MILINF 89,* Försvarets materielverk, Enköping, July 5–9, 1989, Telub AB, Växjö, 1989, pp. 07B2:19-07B2:31.

FOA report B 20078-2.7, Department of Weapon Systems, Effects and Protection, National Defence Research Establishment, Stockholm, 1990.
6. U. Bergsten, J. Schubert, and P. Svensson, "Dezzy—Ett Demonstrationssystem för Analys av Ubåtsunderrättelser," In *Delprojekt Informationssystem inom Huvudprojekt Ubåtsskydd—Slutrapport,* J. Schubert, (Ed.), FOA report A 20046-2.7, Department of Weapon Systems, Effects and Protection, National Defence Research Establishment, Stockholm, 1990, pp. 8–10.
7. J.D. Lowrance and T.D. Garvey, *Evidential Reasoning: An Implementation for Multisensor Integration,* Technical Note 307, Artificial Intelligence Center, Computer Science and Technology Division, SRI International, Menlo Park, CA, 1983.
8. S.A. Lesh, *An Evidential Theory Approach to Judgment-Based Decision Making,* Ph.D. thesis, Dept. of Forestry and Environmental Studies, Duke University, Durham, NC, 1986.

Specifying Nonspecific Evidence

Manuscript

II

Specifying Nonspecific Evidence

Johan Schubert
Division of Information System Technology,
Department of Command and Control Warfare Technology,
National Defence Research Establishment,
S-172 90 Stockholm, Sweden.

In an earlier article [J. Schubert, "On Nonspecific Evidence," *Int. J. Intell. Syst.* **8**(6), 711-725 (1993)] we established within Dempster-Shafer theory a criterion function called the metaconflict function. With this criterion we can partition into subsets a set of evidences with propositions that are weakly specified in the sense that it may be uncertain to which event a proposition is referring. Each subset in the partitioning is representing a separate event. The metaconflict function was derived as the plausibility that the partitioning is correct when viewing the conflict in Dempster's rule within each subset as a newly constructed metalevel evidence with a proposition giving support against the entire partitioning. In this article we extend the results of the previous article. We will not only find the most plausible subset for each piece of evidence as was done in the earlier article. In addition we will specify each nonspecific evidence, in the sense that we find to which events the proposition of the evidence might be referring, by finding the plausibility for every subset that the evidence belongs to the subset. In doing this we will automatically receive an indication on how likely it is that this evidence is actually false. We will then develop a new methodology to exploit these newly specified evidences in a subsequent reasoning process. This will include methods to discount evidences based on their degree of falsity and on their degree of credibility due to a partial specification of affiliation, as well as a refined method to infer the event of each subset.

I. INTRODUCTION

When we are reasoning under uncertainty in an environment of several possible events we may find some evidences that are not only uncertain but may also have propositions that are weakly specified in the sense that it may not be certain to which event a proposition is referring. In addition our own domain knowledge regarding the current number of events may be uncertain. In this situation we must make sure that we do not by mistake combine evidences that are referring to different events.

When the propositions of evidences are weakly specified with respect to which events they are referring, it is impossible to directly judge whether two propositions are referring to the same event. Thus, it is not possible to separate evidences based only on their proposition. Instead we prefer to separate evidences based on their conflicts. Since the conflict measures the lack of compatibility between evidences, and evidences referring to different events tend to be more incompatible than evidences referring to the same event, it is an obvious choice as a distance measure. The idea of using the conflict in Dempster's rule as distance measure between bodies of evidence was first suggested by Lowrance and Garvey.[1]

In an earlier article[2] we established, within the framework of Dempster-Shafer theory,[3-6] a criterion function of overall conflict called the metaconflict function. With this criterion we can partition evidences with weakly specified propositions into subsets, each subset representing a separate event. These events should be handled independently.

To make a separation of evidences possible, every proposition's action part must be supplemented with an event part describing to which event the proposition is referring. If the proposition is written as a conjunction of literals or disjunctions, then one literal or disjunction concerns which event the proposition is referring to. This is the event part. The remainder of the proposition is called the action part. An example from our earlier article illustrates the terminology:

> *Let us consider the burglaries of two bakers' shops at One and Two Baker Street, event* 1 (E_1) *and event* 2 (E_2), *i.e., the number of events is known to be two. One witness hands over an evidence, specific with respect to event, with the proposition: "The burglar at One Baker Street," event part:* E_1, *"was probably brown haired* (B)*," action part:* B. *A second anonymous witness hands over a nonspecific evidence with the proposition: "The burglar at Baker Street," event part:* E_1, E_2, *"might have been red haired* (R)*," action part:* R. *That is, for example:*

evidence 1*:*	*evidence* 2*:*
proposition:	*proposition:*
action part: B	*action part:* R
event part: E_1	*event part:* E_1,E_2
$m(B) = 0.8$	$m(R) = 0.4$
$m(\Theta) = 0.2$	$m(\Theta) = 0.6$

We will have a conflict between two pieces of evidence in the same subset in two different situations. Firstly, we have a conflict if the proposition action parts are conflicting regardless of the proposition event parts since they are presumed to be referring to the same event. Secondly, if the proposition event parts are conflicting then, regardless of the proposition action parts, we have a conflict with the presumption that they are referring to the same event.

The metaconflict used to partition the set of evidences is derived[2] as the plausibility that the partitioning is correct when the conflict in each subset is viewed as a metalevel evidence against the partitioning of the set of evidences, χ, into the subsets, χ_i. We have a simple frame of discernment on the metalevel $\Theta =$

$\{\text{AdP}, \neg\text{AdP}\}$, where AdP is short for "adequate partition", and a basic probability assignment (bpa) from each subset χ_i assigning support to a proposition against the partitioning:

$$m_{\chi_i}(\neg\text{AdP}) \triangleq \text{Conf}(\{e_j | e_j \in \chi_i\}),$$

$$m_{\chi_i}(\Theta) \triangleq 1 - \text{Conf}(\{e_j | e_j \in \chi_i\})$$

where e_j is the jth evidence and $\{e_j | e_j \in \chi_i\}$ is the set of evidences belonging to subset χ_i and $\text{Conf}(\cdot)$ is the conflict, k, in Dempster's rule. Also, we have a bpa concerning the domain resulting from a probability distribution about the number of subsets, E, conflicting with the actual current number of subsets, $\#\chi$. This bpa also assigns support to a proposition against the partitioning:

$$m_D(\neg\text{AdP}) \triangleq \text{Conf}(\{E, \#\chi\}),$$

$$m_D(\Theta) \triangleq 1 - \text{Conf}(\{E, \#\chi\}).$$

The combination of these by Dempster's rule give us the following plausibility of the partitioning:

$$\text{Pls}(\text{AdP}) = (1 - m_D(\neg\text{AdP})) \cdot \prod_{i=1}^{r} (1 - m_{\chi_i}(\neg\text{AdP})).$$

The difference, one minus the plausibility of a partitioning, will be called the metaconflict of the partitioning. The metaconflict function can then be defined as:

DEFINITION.[2] *Let the* metaconflict function,

$$Mcf(r, e_1, e_2, \ldots, e_n) \triangleq 1 - (1 - c_0) \cdot \prod_{i=1}^{r} (1 - c_i),$$

be the conflict against a partitioning of n evidences of the set χ into r disjoint subsets χ_i where

$$c_0 = \sum_{i \neq r} m(E_i)$$

is the conflict between r subsets and propositions about possible different number of subsets, E_i the proposition that there are i subsets, $m(E_i)$ the support for it and

$$c_i = \sum_{\substack{I \\ \cap I = \varnothing}} \prod_{e_j^k \in I} m_j(e_j^k)$$

is the conflict in subset i, where $\cap I$ *is the intersection of all elements in I,* $I = \{e_j^k | e_j \in \chi_i\}$ *is a set of one focal element from the support function of each evidence* e_j *in* χ_i *and* e_j^k *is the* kth *focal element of evidence* e_j.

Thus, $|I| = |\chi_i|$ and

$$|\{I\}| = \prod_{e_j \in \chi_i} |e_j|$$

where $|e_j|$ is the number of focal elements of e_j.

Two theorems are derived[2] to be used in the separation of the set of evidences into subsets by an iterative minimization of the metaconflict function. By using these theorems we are able to reason about the optimal estimate of number of events, when the actual number of events may be uncertain, as well as the optimal partition of nonspecific evidence for any fixed number of events. These two theorems will also be useful in a process for specifying evidences by observing changes in the metaconflict when moving a single piece of evidences between different subsets.

THEOREM 1. *For all j with* $j < r$, *if* $m(E_j) < m(E_r)$ *then* $\min Mcf(r, e_1, e_2, \ldots, e_n) < \min Mcf(j, e_1, e_2, \ldots, e_n)$.

This theorem states that an optimal partitioning for r subsets is always better than the other solutions with fewer than r subsets if the basic probability assignment for r subsets is greater than the basic probability assignment for the fewer subsets.

THEOREM 2. *For all j, if* $\min Mcf(r, e_1, e_2, \ldots, e_n) < \sum_{i \neq j} m(E_i)$ *then* $\min Mcf(r, e_1, e_2, \ldots, e_n) < \min Mcf(j, e_1, e_2, \ldots, e_n)$.

Theorem 2 states that an optimal partitioning for some number of subsets is always better than other solutions for any other number of subsets when the domain part of the metaconflict function is greater than the total metaconflict of our present partitioning.

The methodology to handle and specify nonspecific evidences was developed as a part of a multiple-target tracking algorithm for an anti-submarine intelligence analysis system.[7] In this application a sparse flow of intelligence reports arrives at the analysis system. These reports may originate from several different unconnected sensor systems. The reports carry a proposition about the occurrence of a submarine at a specified time and place, a probability of the truthfulness of the report and may contain additional information such as velocity, direction and type of submarine.

When there are several submarines we want to separate the intelligence reports into subsets according to which submarine they are referring to. We will then analyze the reports for each submarine separately. However, the intelligence reports are never labeled as to which submarine they are referring to. Thus, it is not possible to directly differentiate between two different submarines using two intelligence reports.

Instead we will use the conflict between the propositions of two intelligence reports as a probability that the two reports are referring to different submarines. This probability is the basis for separating intelligence reports into subsets.

The cause of the conflict can be non-firing sensors placed between the positions of the two reports, the required velocity to travel between the positions of the two reports at their respective times in relation to the assumed velocity of the submarines, etc.

In Sec. II we derive a bpa supporting that an evidence is not belonging to a certain subset. This is done by observing the cluster conflict variations when a piece of evidence is moved out from a subset, or when, after that, it is brought into another subset, or by observing the domain conflict variation when it is put into a newly created subset by itself. With this derived bpa we then find the bpa for each evidence and every subset for that evidence. In Sec. III we specify the evidences by combining all bpa's from different subsets regarding one and the same piece of evidence and calculate the plausibility for each subset that the evidence belongs to the subset. In the combination of all bpa's in Sec. III we receive support for a false statement that a piece of evidence does not belong to any of the subsets and can not be placed in a new subset by itself. We discuss how this situation can be handled in Sec. IV. In Sec. V we study the usefulness of the now specified evidences. Obviously, a piece of evidence that can belong to several different subsets is not so useful and should not be allowed to strongly influence a subsequent reasoning process within a subset. We then describe an improved method of finding the event represented by a subset (Sec. VI). Finally, we illustrate the methodology by an example of some "bakers' shops burglaries" and make a comparison of the refined methodology advocated in this article with the simpler approach in our earlier article[2] (Sec. VII).

II. EVIDENCES ABOUT EVIDENCES

A. Evidences from cluster conflict variations

A conflict in a subset can be interpreted as an evidence that there is at least one piece of evidence that does not belong to the subset. Thus, we can refine the basic probability assignment from subset χ_i assigning support to a proposition against the partitioning:

$$m_{\chi_i}(\neg \mathrm{AdP}) = c_i,$$
$$m_{\chi_i}(\Theta) = 1 - c_i$$

to

$$m_{\chi_i}(\exists j.\, e_j \notin \chi_i) = c_i,$$
$$m_{\chi_i}(\Theta) = 1 - c_i.$$

Let us observe one evidence e_q in χ_i. If the evidence e_q is taken out from χ_i the conflict c_i in χ_i decreases to c_i^*.This decrease in conflict $c_i - c_i^*$ can be interpreted as follows: there exists some evidence indicating that e_q does not belong to χ_i,

$$m_{\Delta\chi_i}(e_q \notin \chi_i),$$
$$m_{\Delta\chi_i}(\Theta),$$

and the remainder of the conflict c_i^* is an evidence that there is at least one other evidence e_j, $j \neq q$, that does not belong to χ_i - $\{e_q\}$,

$$m_{\chi_i - \{e_q\}}(\exists j \neq q.\, e_j \notin (\chi_i - \{e_q\})) = c_i^*,$$
$$m_{\chi_i - \{e_q\}}(\Theta) = 1 - c_i^*.$$

We will derive the basic probability number of $e_q \notin \chi_i$, $m_{\Delta\chi_i}(e_q \notin \chi_i)$, by stating that the belief in the proposition that there is at least one piece of evidence that does not belong to χ_i, $\exists j.\, e_j \notin \chi_i$, should be equal no matter whether we base that belief on the original evidence, before e_q is taken out from χ_i, or on a combination of the other two evidences $m_{\Delta\chi_i}(e_q \notin \chi_i)$ and $m_{\chi_i - \{e_q\}}(\exists j \neq q.\, e_j \notin (\chi_i - \{e_q\}))$, after e_q is taken out from χ_i, i.e.

$$\mathrm{Bel}_{\chi_i}(\exists j.\, e_j \notin \chi_i) = \mathrm{Bel}_{\Delta\chi_i \oplus (\chi_i - \{e_q\})}(\exists j.\, e_j \notin \chi_i).$$

We may rewrite the original proposition

$$\exists j.\, e_j \notin \chi_i$$

as

$$(\exists j \neq q.\, e_j \notin \chi_i) \vee (e_q \notin \chi_i)$$

and as

$$(\exists j \neq q.\, e_j \notin (\chi_i - \{e_q\})) \vee (e_q \notin \chi_i).$$

Then, we have on the one hand, before e_q is taken out from χ_i,

$$\mathrm{Bel}_{\chi_i}(\exists j.\, e_j \notin \chi_i) = m_{\chi_i}(\exists j.\, e_j \notin \chi_i) = c_i,$$

and on the other hand, if the evidence that there is at least one other evidence e_j, $j \neq q$, that does not belong to χ_i - $\{e_q\}$ is fused with the evidence that e_q does not belong to χ_i, Figure 1, we may also calculate the belief in $\exists j. e_j \notin \chi_i$ as

$$\begin{aligned}
&\mathrm{Bel}_{\Delta\chi_i \oplus (\chi_i - \{e_q\})}(\exists j. e_j \notin \chi_i) \\
&= \mathrm{Bel}_{\Delta\chi_i \oplus (\chi_i - \{e_q\})}((\exists j \neq q. e_j \notin (\chi_i - \{e_q\})) \vee (e_q \notin \chi_i)) \\
&= \sum_{X \subseteq ((\exists j \neq q. e_j \notin (\chi_i - \{e_q\})) \vee (e_q \notin \chi_i))} m_{\Delta\chi_i \oplus (\chi_i - \{e_q\})}(X) \\
&= m_{\Delta\chi_i \oplus (\chi_i - \{e_q\})}((\exists j \neq q. e_j \notin (\chi_i - \{e_q\})) \wedge (e_q \notin \chi_i)) \\
&\quad + m_{\Delta\chi_i \oplus (\chi_i - \{e_q\})}(\exists j \neq q. e_j \notin (\chi_i - \{e_q\})) + m_{\Delta\chi_i \oplus (\chi_i - \{e_q\})}(e_q \notin \chi_i) \\
&= m_{\chi_i - \{e_q\}}(\exists j \neq q. e_j \notin (\chi_i - \{e_q\})) \cdot m_{\Delta\chi_i}(e_q \notin \chi_i) \\
&\quad + m_{\chi_i - \{e_q\}}(\exists j \neq q. e_j \notin (\chi_i - \{e_q\})) \cdot [1 - m_{\Delta\chi_i}(e_q \notin \chi_i)] \\
&\quad + m_{\Delta\chi_i}(e_q \notin \chi_i) \cdot [1 - m_{\chi_i - \{e_q\}}(\exists j \neq q. e_j \notin (\chi_i - \{e_q\}))] \\
&= m_{\chi_i - \{e_q\}}(\exists j \neq q. e_j \notin (\chi_i - \{e_q\})) \\
&\quad + m_{\Delta\chi_i}(e_q \notin \chi_i) \cdot [1 - m_{\chi_i - \{e_q\}}(\exists j \neq q. e_j \notin (\chi_i - \{e_q\}))] \\
&= c_i^* + m_{\Delta\chi_i}(e_q \notin \chi_i) \cdot [1 - c_i^*] .
\end{aligned}$$

Thus, we have derived the evidence that e_q does not belong to χ_i from the variations in cluster conflict when e_q was taken out from χ_i:

$$m_{\Delta\chi_i}(e_q \notin \chi_i) = \frac{c_i - c_i^*}{1 - c_i^*},$$

$$m_{\Delta\chi_i}(\Theta) = 1 - \frac{c_i - c_i^*}{1 - c_i^*} = \frac{1 - c_i}{1 - c_i^*}.$$

If e_q then is brought into another subset χ_k its conflict will increase from c_k to c_k^* where

	$\exists j \neq q. e_j \notin (\chi_i - \{e_q\})$	Θ
$e_q \notin \chi_i$	$(\exists j \neq q. e_j \notin (\chi_i - \{e_q\})) \wedge (e_q \notin \chi_i)$	$e_q \notin \chi_i$
Θ	$\exists j \neq q. e_j \notin (\chi_i - \{e_q\})$	Θ

Figure 1. Combining $\mathrm{Bel}_{\chi_i - \{e_q\}}$ and $\mathrm{Bel}_{\Delta\chi_i}$.

$$\forall k \neq i. m_{\chi_k}(\exists j \neq q. e_j \notin \chi_k) = c_k,$$
$$\forall k \neq i. m_{\chi_k}(\Theta) = 1 - c_k.$$

and

$$\forall k \neq i. m_{\chi_k + \{e_q\}}(\exists j. e_j \notin (\chi_k + \{e_q\})) = c_k^*,$$
$$\forall k \neq i. m_{\chi_k + \{e_q\}}(\Theta) = 1 - c_k^*.$$

Thus we will also have evidences regarding every other subset χ_k.

The increase in conflict when e_q is brought into χ_k is interpreted as if there exists some evidence indicating that e_q does not belong to $\chi_k + \{e_q\}$, i.e.

$$m_{\Delta\chi_k}(e_q \notin (\chi_k + \{e_q\})),$$
$$m_{\Delta\chi_k}(\Theta).$$

Similarly to the last case, we may this time rewrite the new proposition

$$\exists j. e_j \notin (\chi_k + \{e_q\})$$

as

$$(\exists j \neq q. e_j \notin (\chi_k + \{e_q\})) \vee (e_q \notin (\chi_k + \{e_q\}))$$

and as

$$(\exists j \neq q. e_j \notin \chi_k) \vee (e_q \notin (\chi_k + \{e_q\}))$$

Reasoning in the same way as above, we state that

$$\forall k \neq i. \mathrm{Bel}_{\chi_k + \{e_q\}}(\exists j. e_j \notin (\chi_k + \{e_q\})) = \mathrm{Bel}_{\Delta\chi_k \oplus \chi_k}(\exists j. e_j \notin (\chi_k + \{e_q\}))$$

and find that on the one hand, after e_q is brought into χ_k

$$\forall k \neq i. \mathrm{Bel}_{\chi_k + \{e_q\}}(\exists j. e_j \notin (\chi_k + \{e_q\})) = m_{\chi_k + \{e_q\}}(\exists j. e_j \notin (\chi_k + \{e_q\})) = c_k^*$$

and on the other hand,

$$\forall k \neq i.\,\mathrm{Bel}_{\Delta\chi_k \oplus \chi_k}(\exists j.\, e_j \notin (\chi_k + \{e_q\}))$$
$$= \mathrm{Bel}_{\Delta\chi_k \oplus \chi_k}((\exists j \neq q.\, e_j \notin \chi_k) \vee (e_q \notin (\chi_k + \{e_q\})))$$
$$= \sum_{X \subseteq ((\exists j \neq q.\, e_j \notin \chi_k) \vee (e_q \notin (\chi_k + \{e_q\})))} m_{\Delta\chi_k \oplus \chi_k}(X)$$
$$= m_{\Delta\chi_k \oplus \chi_k}((\exists j \neq q.\, e_j \notin \chi_k) \wedge (e_q \notin (\chi_k + \{e_q\}))) + m_{\Delta\chi_k \oplus \chi_k}(\exists j \neq q.\, e_j \notin \chi_k)$$
$$+ m_{\Delta\chi_k \oplus \chi_k}(e_q \notin (\chi_k + \{e_q\}))$$
$$= m_{\chi_k}(\exists j \neq q.\, e_j \notin \chi_k) \cdot m_{\Delta\chi_k}(e_q \notin (\chi_k + e_q))$$
$$+ m_{\chi_k}(\exists j \neq q.\, e_j \notin \chi_k) \cdot [1 - m_{\Delta\chi_k}(e_q \notin (\chi_k + \{e_q\}))]$$
$$+ m_{\Delta\chi_k}(e_q \notin (\chi_k + \{e_q\})) \cdot [1 - m_{\chi_k}(\exists j \neq q.\, e_j \notin \chi_k)]$$
$$= m_{\chi_k}(\exists j \neq q.\, e_j \notin \chi_k) + m_{\Delta\chi_k}(e_q \notin (\chi_k + \{e_q\})) \cdot [1 - m_{\chi_k}(\exists j \neq q.\, e_j \notin \chi_k)]$$
$$= c_k + m_{\Delta\chi_k}(e_q \notin (\chi_k + \{e_q\})) \cdot [1 - c_k].$$

Thus, we have then derived an evidence regarding each subset $\chi_k + \{e_q\}$, $k \neq i$, that e_q does not belong to the subset;

$$\forall k \neq i.\, m_{\Delta\chi_k}(e_q \notin (\chi_k + \{e_q\})) = \frac{c_k^* - c_k}{1 - c_k},$$
$$\forall k \neq i.\, m_{\Delta\chi_k}(\Theta) = 1 - \frac{c_k^* - c_k}{1 - c_k} = \frac{1 - c_k^*}{1 - c_k}.$$

B. Evidences from domain conflict variations

Since an evidence from domain conflict is an evidence against the entire partitioning it is less specific than an evidence from cluster conflict. We will interpret the domain conflict as evidence that there exists at least one piece of evidence that does not belong to any of the n first subsets, or if this particular evidence is in a subset by itself as evidence that it belongs to some of the other n-1 subsets. This would indicating that the number of subsets is incorrect.

We will now study any changes in the domain conflict when we take out an evidence e_q from subset χ_i.

When $|\chi_i| > 1$ we may not only put the evidence e_q that we have taken out from χ_i into another already existing subset, we may also put e_q into a new subset χ_{n+1} by itself, assuming there are n subsets, i.e. $\chi = \{\chi_1, \ldots, \chi_n\}$. This will change the domain conflict, c_0. Since the current partition minimizes the metaconflict function, we know that when the number of subsets increase we will get an increase in total conflict and Theorem 1 says that we will get a decrease in the nondomain part of the metaconflict function. Thus, we know that we must get an increase in the domain conflict. This increase in domain conflict is an indication that e_q does not belong to an additional subset χ_{n+1}.

Another way to receive an evidence from the domain conflict is if e_q is moved out from χ_i when $|\chi_i| = 1$. If e_q is in a subset χ_i by itself and moved from χ_i to another already existing subset we may get either an increase or decrease in domain conflict. This is because both the total conflict and the nondomain part of the metaconflict function increases. Thus, we have two different situations in this case. If the domain conflict decreases when we move e_q out from χ_i this is interpreted as an evidence that e_q does not belongs to χ_i, but if we receive an increase in domain conflict we will interpret this as an evidence that e_q does belong to χ_i.

We choose to adopt a metarepresentation consisting of three individual representations for the domain conflict. The first representation interprets the domain conflict as an evidence that there is at least one piece of evidence that does not belong to any of the subsets,

$$\exists j \forall k . e_j \notin \chi_k .$$

The second representation interprets the domain conflict as an evidence that there is at least one subset to which no evidences belongs,

$$\exists k \forall j . e_j \notin \chi_k ,$$

and the third as an evidence that there is either at least one piece of evidence that does not belong to any of the subsets or there is at least one subset to which no evidences belongs, but not both at the same time,

$$\begin{gathered}[(\exists j \forall k . e_j \notin \chi_k) \wedge (\neg \exists k \forall j . e_j \notin \chi_k)] \\ \vee [(\neg \exists j \forall k . e_j \notin \chi_k) \wedge (\exists k \forall j . e_j \notin \chi_k)] .\end{gathered}$$

Each representation has its own characteristics. The first and the second are consistent with a situation where the domain conflict increases when the number of subsets increase and decreases when the number of subsets decrease. The third representation behaves in the opposite way. The three representations above correspond to these three different situations when the number of subsets is changed.

The first representation corresponds to the situation when one evidence e_q belongs to a subset χ_i, $|\chi_i| > 1$, and the evidence is moved from χ_i to χ_{n+1}. This increases both the domain conflict and the number of subsets. The second representation is not in accordance with this situation and the third is not even consistent with the situation. The second representation corresponds to the situation when one evidence e_q belongs to a subset χ_i, $|\chi_i| = 1$, and the evidence is moved from χ_i to one of the other n-1 subsets while we receive a decrease in domain conflict. Here, the first representation is not in accordance with this situation and the third is not consistent. The third representation corresponds to the last situation when one evidence e_q belongs to a subset χ_i, $|\chi_i| = 1$, and the evidence is moved from χ_i to one of the other n-1 subsets while we receive an increase in domain conflict. In this situation the first and second representations are not consistent.

Thus, the actual representation to be used can be chosen from the metarepresentation by the current situation. Let us now see what can be derived about our evidence of interest, e_q.

1. *When* $e_q \in \chi_i$ *and* $|\chi_i| > 1$

Let us analyze the case where we move e_q from χ_i to χ_{n+1}. The domain conflict before e_q is moved to χ_{n+1} is interpreted as

$$\exists j \forall k \neq n+1 . e_j \notin \chi_k .$$

When $e_q \in \chi_i$ and $|\chi_i| > 1$ we may move out e_q from χ_i without changing the domain conflict, but we will get an increase in the domain conflict if we move e_q to a subset by itself; χ_{n+1}.

If e_q is taken out from χ_i, without being moved to χ_{n+1}, and is temporarily disregarded from the analysis we will still have n subsets in this new situation since $|\chi_i|$ was greater than one. The domain conflict, which is unchanged by this removal and equal to c_0, may be interpreted in this case as

$$\exists j \neq q \forall k \neq n+1 . e_j \notin \chi_k .$$

Thus, when e_q is taken out from χ_i, we can refine the bpa regarding the domain conflict to

$$m_\chi(\exists j \neq q \forall k \neq n+1 . e_j \notin \chi_k) = c_0 ,$$
$$m_\chi(\Theta) = 1 - c_0 .$$

If e_q then is moved to χ_{n+1} the increase in domain conflict is an evidence that e_q does not belong to χ_{n+1},

$$m_{\Delta\chi}(e_q \notin \chi_{n+1}),$$
$$m_{\Delta\chi}(\Theta),$$

and the new domain conflict that we receive indicate that there must now be at least one piece of evidence that does not belong to any of the n+1 subsets,

$$m_{\chi + \{\chi_{n+1}\}}(\exists j \forall k . e_j \notin \chi_k) = c_0^* ,$$
$$m_{\chi + \{\chi_{n+1}\}}(\Theta) = 1 - c_0^* .$$

We will derive $m_{\Delta\chi}(e_q \notin \chi_{n+1})$ by stating that

$$\mathrm{Bel}_{\chi + \{\chi_{n+1}\}}(\exists j \forall k. e_j \notin \chi_k) = \mathrm{Bel}_{\Delta\chi \oplus \chi}(\exists j \forall k. e_j \notin \chi_k).$$

After e_q is moved to χ_{n+1}, we may rewrite the proposition

$$\exists j \forall k. e_j \notin \chi_k$$

as

$$[\exists j \neq q(((\forall k \neq n+1. e_j \notin \chi_k) \wedge (e_j \notin \chi_{n+1}))]$$
$$\vee [((\forall k \neq n+1. e_q \notin \chi_k) \wedge (e_q \notin \chi_{n+1})]$$

and as

$$(\exists j \neq q \forall k \neq n+1. e_j \notin \chi_k) \vee (e_q \notin \chi_{n+1})$$

since we know it is true that $e_j \notin \chi_{n+1}$ for some other evidences than e_q and it is true that $e_q \notin \chi_k$ for all other subsets than χ_{n+1}, since e_q is now in χ_{n+1}.

Then, on the one hand we have

$$\mathrm{Bel}_{\chi + \{\chi_{n+1}\}}(\exists j \forall k. e_j \notin \chi_k) = m_{\chi + \{\chi_{n+1}\}}(\exists j \forall k. e_j \notin \chi_k) = c_0^*$$

and on the other hand we can calculate

$$\mathrm{Bel}_{\Delta\chi \oplus \chi}(\exists j \forall k. e_j \notin \chi_k) = \mathrm{Bel}_{\Delta\chi \oplus \chi}((\exists j \neq q \forall k \neq n+1. e_j \notin \chi_k) \vee (e_q \notin \chi_{n+1}))$$
$$= \sum_{X \subseteq ((\exists j \neq q \forall k \neq n+1. e_j \notin \chi_k) \vee (e_q \notin \chi_{n+1}))} m_{\Delta\chi \oplus \chi}(X)$$
$$= m_{\Delta\chi \oplus \chi}((\exists j \neq q \forall k \neq n+1. e_j \notin \chi_k) \wedge (e_q \notin \chi_{n+1}))$$
$$m_{\Delta\chi \oplus \chi}(\exists j \neq q \forall k \neq n+1. e_j \notin \chi_k) + m_{\Delta\chi \oplus \chi}(e_q \notin \chi_{n+1})$$
$$= m_\chi(\exists j \neq q \forall k \neq n+1. e_j \notin \chi_k) \cdot m_{\Delta\chi}(e_q \notin \chi_{n+1})$$
$$+ m_\chi(\exists j \neq q \forall k \neq n+1. e_j \notin \chi_k) \cdot [1 - m_{\Delta\chi}(e_q \notin \chi_{n+1})]$$
$$+ m_{\Delta\chi}(e_q \notin \chi_{n+1}) \cdot [1 - m_\chi(\exists j \neq q \forall k \neq n+1. e_j \notin \chi_k)]$$
$$= m_\chi(\exists j \neq q \forall k \neq n+1. e_j \notin \chi_k) + m_{\Delta\chi}(e_q \notin \chi_{n+1}) \cdot [1 - m_\chi(\exists j \neq q \forall k \neq n+1. e_j \notin \chi_k)]$$
$$= c_0 + m_{\Delta\chi}(e_q \notin \chi_{n+1}) \cdot [1 - c_0].$$

Thus, we get

$$m_{\Delta\chi}(e_q \notin \chi_{n+1}) = \frac{c_0^* - c_0}{1 - c_0}$$

an evidence indicating that e_q does not belong to χ_{n+1}.

2. When $e_q \in \chi_i$, $|\chi_i| = 1$ and $c_0 > c_0^*$

Let us now study the situation when e_q is in a subset by itself. Before e_q is moved out from χ_i,

$$\exists k \forall j.\, e_j \notin \chi_k$$

represents the domain conflict with the n subsets. Thus, we may refine the bpa regarding domain conflict in this situation to

$$m_\chi(\exists k \forall j.\, e_j \notin \chi_k) = c_0,$$
$$m_\chi(\Theta) = 1 - c_0.$$

If we take out e_q from χ_i, without moving it to any already existing subset, and temporarily disregard it from the analysis we have n-1 remaining subsets and get a decrease in the domain conflict. This decrease in domain conflict is interpreted as an evidence that e_q does not belong to χ_i,

$$m_{\Delta\chi}(e_q \notin \chi_i),$$
$$m_{\Delta\chi}(\Theta).$$

The remaining domain conflict indicate that there is now at least one other subset χ_k, $k \neq i$, that does not contain any evidences e_j, $j \neq q$,

$$m_{\chi - \{\chi_i\}}(\exists k \neq i \forall j \neq q.\, e_j \notin \chi_k) = c_0^*,$$
$$m_{\chi - \{\chi_i\}}(\Theta) = 1 - c_0^*.$$

As before, we will derive $m_{\Delta\chi}(e_q \notin \chi_i)$ by stating that

$$\mathrm{Bel}_\chi(\exists k \forall j.\, e_j \notin \chi_k) = \mathrm{Bel}_{\Delta\chi \oplus (\chi - \{\chi_i\})}(\exists k \forall j.\, e_j \notin \chi_k).$$

Before e_q is taken out from χ_i we may rewrite the proposition

$$\exists k \forall j.\, e_j \notin \chi_k$$

as

$$[\exists k \neq i((\forall j \neq q.\, e_j \notin \chi_k) \wedge (e_q \notin \chi_k))] \vee [(\forall j \neq q.\, e_j \notin \chi_i) \wedge (e_q \notin \chi_i)]$$

and as

$$(\exists k \neq i \forall j \neq q.\, e_j \notin \chi_k) \vee (e_q \notin \chi_i).$$

since $e_q \notin \chi_k$ for some $k \neq i$ and $e_j \notin \chi_i$ for all other evidences than e_q, since e_q is still in χ_i.

Similarly to the previous case we have

$$\mathrm{Bel}_{\chi}(\exists k \forall j . e_j \notin \chi_k) = m_{\chi}(\exists k \forall j . e_j \notin \chi_k) = c_0$$

and may calculate

$$\begin{aligned}
&\mathrm{Bel}_{\Delta\chi \oplus (\chi - \{\chi_i\})}(\exists k \forall j . e_j \notin \chi_k) \\
&= \mathrm{Bel}_{\Delta\chi \oplus (\chi - \{\chi_i\})}((\exists k \neq i \forall j \neq q . e_j \notin \chi_k) \vee (e_q \notin \chi_i)) \\
&= \sum_{X \subseteq ((\exists k \neq i \forall j \neq q . e_j \notin \chi_k) \vee (e_q \notin \chi_i))} m_{\Delta\chi \oplus (\chi - \{\chi_i\})}(X) \\
&= m_{\Delta\chi \oplus (\chi - \{\chi_i\})}((\exists k \neq i \forall j \neq q . e_j \notin \chi_k) \wedge (e_q \notin \chi_i)) \\
&\quad + m_{\Delta\chi \oplus (\chi - \{\chi_i\})}(\exists k \neq i \forall j \neq q . e_j \notin \chi_k) + m_{\Delta\chi \oplus (\chi - \{\chi_i\})}(e_q \notin \chi_i) \\
&= m_{\chi - \{\chi_i\}}(\exists k \neq i \forall j \neq q . e_j \notin \chi_k) \cdot m_{\Delta\chi}(e_q \notin \chi_i) \\
&\quad + m_{\chi - \{\chi_i\}}(\exists k \neq i \forall j \neq q . e_j \notin \chi_k) \cdot [1 - m_{\Delta\chi}(e_q \notin \chi_i)] \\
&\quad + m_{\Delta\chi}(e_q \notin \chi_i) \cdot [1 - m_{\chi - \{\chi_i\}}(\exists k \neq i \forall j \neq q . e_j \notin \chi_k)] \\
&= m_{\chi - \{\chi_i\}}(\exists k \neq i \forall j \neq q . e_j \notin \chi_k) + m_{\Delta\chi}(e_q \notin \chi_i) \cdot [1 - m_{\chi - \{\chi_i\}}(\exists k \neq i \forall j \neq q . e_j \notin \chi_k)] \\
&= c_0^* + m_{\Delta\chi}(e_q \notin \chi_i) \cdot [1 - c_0^*] .
\end{aligned}$$

Thus, when $c_0 > c_0^*$ we have

$$m_{\Delta\chi}(e_q \notin \chi_i) = \frac{c_0 - c_0^*}{1 - c_0^*} .$$

3. *When* $e_q \in \chi_i$, $|\chi_i| = 1$ *and* $c_0 < c_0^*$

A completely different situation occurs when $c_0 < c_0^*$. This is the case when we choose to represent the domain conflict as an exclusive-OR of two propositions,

$$\begin{aligned}
&[(\exists j \forall k . e_j \notin \chi_k) \wedge (\neg \exists k \forall j . e_j \notin \chi_k)] \\
&\vee [(\neg \exists j \forall k . e_j \notin \chi_k) \wedge (\exists k \forall j . e_j \notin \chi_k)] ,
\end{aligned}$$

one stating that there is at least one piece of evidence that does not belong to any of the n subsets and the other one stating that there is at least one subset that does not contain any evidences. Thus, we can refine the bpa regarding domain conflict to

$$m_\chi\left(\begin{array}{c}[(\exists j\forall k.\, e_j \notin \chi_k) \wedge (\neg\exists k\forall j.\, e_j \notin \chi_k)] \\ \vee\, [(\neg\exists j\forall k.\, e_j \notin \chi_k) \wedge (\exists k\forall j.\, e_j \notin \chi_k)]\end{array}\right) = c_0,$$

$$m_\chi(\Theta) = 1 - c_0 .$$

If we take out e_q from χ_i as in the previous case, without moving it to any already existing subset, and temporarily disregard it from the analysis we have n-1 remaining subsets since $|\chi_i|$ was equal to one and get an increase in the domain conflict. This increase in domain conflict will be interpreted as an evidence that e_q belongs to χ_i,

$$m_{\Delta\chi}(e_q \in \chi_i),$$

$$m_{\Delta\chi}(\Theta).$$

The remaining domain conflict indicate that there is now at least one other subset χ_k, $k \neq i$, that does not contain any evidences e_j, $j \neq q$, or that there is at least one evidence e_j, $j \neq q$, that does not belong to any of the n-1 subsets χ_k, $k \neq i$,

$$m_{\chi - \{\chi_i\}}\left(\begin{array}{c}((\exists j \neq q \forall k \neq i.\, e_j \notin \chi_k) \wedge (\neg\exists k \neq i \forall j \neq q.\, e_j \in \chi_k)) \\ \vee\, ((\neg\exists j \neq q \forall k \neq i.\, e_j \in \chi_k) \wedge (\exists k \neq i \forall j \neq q.\, e_j \notin \chi_k))\end{array}\right) = c_0^*,$$

$$m_{\chi - \{\chi_i\}}(\Theta) = 1 - c_0^* .$$

We will derive $m_{\Delta\chi}(\cdot)$ by stating that

$$\mathrm{Bel}_\chi\left(\begin{array}{c}[(\exists j\forall k.\, e_j \notin \chi_k) \wedge (\neg\exists k\forall j.\, e_j \notin \chi_k)] \\ \vee\, [(\neg\exists j\forall k.\, e_j \notin \chi_k) \wedge (\exists k\forall j.\, e_j \notin \chi_k)]\end{array}\right)$$

$$= \mathrm{Bel}_{\Delta\chi \oplus (\chi - \{\chi_i\})}\left(\begin{array}{c}[(\exists j\forall k.\, e_j \notin \chi_k) \wedge (\neg\exists k\forall j.\, e_j \notin \chi_k)] \\ \vee\, [(\neg\exists j\forall k.\, e_j \notin \chi_k) \wedge (\exists k\forall j.\, e_j \notin \chi_k)]\end{array}\right).$$

The domain conflict, represented by

$$\begin{array}{c}[(\exists j\forall k.\, e_j \notin \chi_k) \wedge (\neg\exists k\forall j.\, e_j \notin \chi_k)] \\ \vee\, [(\neg\exists j\forall k.\, e_j \notin \chi_k) \wedge (\exists k\forall j.\, e_j \notin \chi_k)],\end{array}$$

can be rewritten as

$$\begin{array}{c}[((\exists j \neq q \forall k \neq i.\, e_j \notin \chi_k) \vee (e_q \notin \chi_i)) \wedge ((\forall k \neq i \exists j \neq q.\, e_j \in \chi_k) \wedge (e_q \in \chi_i))] \\ \vee\, [((\forall j \neq q \exists k \neq i.\, e_j \in \chi_k) \wedge (e_q \in \chi_i)) \wedge ((\exists k \neq i \forall j \neq q.\, e_j \notin \chi_k) \vee (e_q \notin \chi_i))],\end{array}$$

by using the simplifications of the previous two sections and further simplified as

$$((\exists j \neq q \forall k \neq i . e_j \notin \chi_k) \wedge (\forall k \neq i \exists j \neq q . e_j \in \chi_k) \wedge (e_q \in \chi_i))$$
$$\vee ((\forall j \neq q \exists k \neq i . e_j \in \chi_k) \wedge (\exists k \neq i \forall j \neq q . e_j \notin \chi_k) \vee (e_q \in \chi_i)),$$

and finally restated as

$$[\ ((\exists j \neq q \forall k \neq i . e_j \notin \chi_k) \wedge (\neg\exists k \neq i \forall j \neq q . e_j \in \chi_k))$$
$$\vee ((\neg\exists j \neq q \forall k \neq i . e_j \in \chi_k) \wedge (\exists k \neq i \forall j \neq q . e_j \notin \chi_k))\ ] \wedge (e_q \in \chi_i)$$

where the first part is interpreted as the domain conflict before e_q is moved from χ_i and the second a proposition stating that e_q belongs to χ_i.

Then, on the one hand we have

$$\mathrm{Bel}_\chi\begin{pmatrix} [(\exists j \forall k . e_j \notin \chi_k) \wedge (\neg\exists k \forall j . e_j \notin \chi_k)] \\ \vee [(\neg\exists j \forall k . e_j \notin \chi_k) \wedge (\exists k \forall j . e_j \notin \chi_k)] \end{pmatrix}$$
$$= m_\chi\begin{pmatrix} [(\exists j \forall k . e_j \notin \chi_k) \wedge (\neg\exists k \forall j . e_j \notin \chi_k)] \\ \vee [(\neg\exists j \forall k . e_j \notin \chi_k) \wedge (\exists k \forall j . e_j \notin \chi_k)] \end{pmatrix} = c_0$$

and on the other hand we can calculate

$$\mathrm{Bel}_{\Delta\chi \oplus (\chi - \{\chi_i\})}\begin{pmatrix} [(\exists j \forall k . e_j \notin \chi_k) \wedge (\neg\exists k \forall j . e_j \notin \chi_k)] \\ \vee [(\neg\exists j \forall k . e_j \notin \chi_k) \wedge (\exists k \forall j . e_j \notin \chi_k)] \end{pmatrix}$$
$$= \mathrm{Bel}_{\Delta\chi \oplus (\chi - \{\chi_i\})}\begin{pmatrix} [\ ((\exists j \neq q \forall k \neq i . e_j \notin \chi_k) \wedge (\neg\exists k \neq i \forall j \neq q . e_j \in \chi_k)) \\ \vee ((\neg\exists j \neq q \forall k \neq i . e_j \in \chi_k) \wedge (\exists k \neq i \forall j \neq q . e_j \notin \chi_k))\ ] \\ \wedge (e_q \in \chi_i) \end{pmatrix}$$
$$= \sum_{X \subseteq \begin{pmatrix} [\ ((\exists j \neq q \forall k \neq i . e_j \notin \chi_k) \wedge (\neg\exists k \neq i \forall j \neq q . e_j \in \chi_k)) \\ \vee ((\neg\exists j \neq q \forall k \neq i . e_j \in \chi_k) \wedge (\exists k \neq i \forall j \neq q . e_j \notin \chi_k))\ ] \wedge (e_q \in \chi_i) \end{pmatrix}} m_{\Delta\chi \oplus (\chi - \{\chi_i\})}(X)$$
$$= m_{\Delta\chi \oplus (\chi - \{\chi_i\})}\begin{pmatrix} [\ ((\exists j \neq q \forall k \neq i . e_j \notin \chi_k) \wedge (\neg\exists k \neq i \forall j \neq q . e_j \in \chi_k)) \\ \vee ((\neg\exists j \neq q \forall k \neq i . e_j \in \chi_k) \wedge (\exists k \neq i \forall j \neq q . e_j \notin \chi_k))\ ] \\ \wedge (e_q \in \chi_i) \end{pmatrix}$$
$$= m_{\chi - \{\chi_i\}}\begin{pmatrix} [\ ((\exists j \neq q \forall k \neq i . e_j \notin \chi_k) \wedge (\neg\exists k \neq i \forall j \neq q . e_j \in \chi_k)) \\ \vee ((\neg\exists j \neq q \forall k \neq i . e_j \in \chi_k) \wedge (\exists k \neq i \forall j \neq q . e_j \notin \chi_k))\ ] \end{pmatrix} \cdot m_{\Delta\chi}(e_q \in \chi_i)$$
$$= c_0^* \cdot m_{\Delta\chi}(e_q \in \chi_i).$$

Thus, if $c_0 < c_0^*$ then

$$m_{\Delta\chi}(e_q \in \chi_i) = \frac{c_0}{c_0^*}$$

is derived as our evidence.

C. Summary of evidences

In summary we have the following evidences

$$\forall i, e_q \in \chi_i . m(e_q \notin \chi_j) = \begin{cases} \dfrac{c_0^* - c_0}{1 - c_0}, j = n+1, |\chi_i| > 1 \\ \dfrac{c_i - c_i^*}{1 - c_i^*}, j = i, |\chi_i| > 1 \\ \dfrac{c_0 - c_0^*}{1 - c_0^*}, j = i, |\chi_i| = 1, c_0 > c_0^* \\ \dfrac{c_j^* - c_j}{1 - c_j}, \text{otherwise} \end{cases}$$

and

$$\forall i, e_q \in \chi_i . m(e_q \in \chi_i) = \frac{c_0}{c_0^*}, |\chi_i| = 1, c_0 < c_0^* .$$

III. SPECIFYING EVIDENCES

We may now specify any original evidence by combining the evidences from conflict variations regarding this particular evidence. Then we may calculate the belief and plausibility for each subset that this particular evidence belongs to the subset. The belief that this particular evidence belongs to a subset will be zero, except when $e_q \in \chi_i$, $|\chi_i| = 1$ and $c_0 < c_0^*$, since every proposition regarding this evidence then states that the evidence does not belong to some subset.

In combining all evidences regarding an original piece of evidence we may receive support for a proposition stating that this piece of evidence does not belong to any of the subsets and can not be put into a subset by itself. Since this is impossible, the statement is false and its support is the conflict in Dempster's rule. The statement that the evidence does not belong anywhere implies that the evidence is false. Thus, we may interpret the conflict as support for the evidence being false.

A. Combining evidences about evidences

1. When $e_q \in \chi_i$ and $|\chi_i| > 1$

Let us assume that an evidence, e_q, is in χ_i and $|\chi_i| > 1$. When we combine all evidences regarding e_q this results in a new basic probability assignment with

$$\forall\chi^* . m^*(e_q \notin (\ \vee\chi^*)) = \prod_{\chi_j \in \chi^*} m(e_q \notin \chi_j) \cdot \prod_{\chi_j \in (\chi - \chi^*)} [1 - m(e_q \notin \chi_j)]$$

where $\chi^* \in 2^{\chi}$ and $\chi = \{\chi_1, \ldots, \chi_{n+1}\}$.

From the new bpa we can calculate the conflict. The only statement that is false is the statement that $e_q \notin (\ \vee\chi)$, i.e. that $\forall j . e_q \notin \chi_j$.

Thus, the conflict becomes

$$k = m^*(e_q \notin (\ \vee\chi)) = \prod_{j=1}^{n+1} m(e_q \notin \chi_j) = \frac{c_0^* - c_0}{1 - c_0} \cdot \prod_{j=1}^{n} \frac{c_j^* - c_j}{1 - c_j}.$$

When calculating belief and plausibility that e_q belongs to some subset other than χ_{n+1} we have

$$\forall k \neq n+1 . \mathrm{Bel}(e_q \in \chi_k) = 1 - \sum_{X \subseteq (e_q \in \chi_k)} m(X) = 0$$

and

$$\forall k \neq n+1 . \mathrm{Pls}(e_q \in \chi_k) = 1 - \mathrm{Bel}(e_q \notin \chi_k) = 1 - \sum_{X \subseteq (e_q \notin \chi_k)} m(X)$$

$$= 1 - \frac{1}{1-k} \cdot m(e_q \notin \chi_k) \cdot \left[1 - \prod_{\substack{j=1 \\ \neq k}}^{n+1} m(e_q \notin \chi_j) \right]$$

$$= 1 - \frac{m(e_q \notin \chi_k) - \prod_{j=1}^{n+1} m(e_q \notin \chi_j)}{1 - \prod_{j=1}^{n+1} m(e_q \notin \chi_j)} = \frac{1 - m(e_q \notin \chi_k)}{1 - \prod_{j=1}^{n+1} m(e_q \notin \chi_j)}$$

$$= \frac{1 - \frac{c_k^* - c_k}{1 - c_k}}{1 - \frac{c_0^* - c_0}{1 - c_0} \cdot \prod_{j=1}^{n} \frac{c_j^* - c_j}{1 - c_j}}$$

while for the subset χ_{n+1} we have

$$\mathrm{Bel}(e_q \in \chi_{n+1}) = 1 - \sum_{X \subseteq (e_q \in \chi_{n+1})} m(X) = 0$$

and

$$\mathrm{Pls}(e_q \in \chi_{n+1}) = 1 - \mathrm{Bel}(e_q \notin \chi_{n+1}) = 1 - \sum_{X \subseteq (e_q \notin \chi_{n+1})} m(X)$$

$$= 1 - \frac{1}{1-k} \cdot m(e_q \notin \chi_{n+1}) \cdot \left[1 - \prod_{j=1}^{n} m(e_q \notin \chi_j)\right]$$

$$= 1 - \frac{m(e_q \notin \chi_{n+1}) - \prod_{j=1}^{n+1} m(e_q \notin \chi_j)}{1 - \prod_{j=1}^{n+1} m(e_q \notin \chi_j)} = \frac{1 - m(e_q \notin \chi_{n+1})}{1 - \prod_{j=1}^{n+1} m(e_q \notin \chi_j)}$$

$$= \frac{1 - \frac{c_0^* - c_0}{1 - c_0}}{1 - \frac{c_0^* - c_0}{1 - c_0} \cdot \prod_{j=1}^{n} \frac{c_j^* - c_j}{1 - c_j}}.$$

2. *When* $e_q \in \chi_i$, $|\chi_i| = 1$ *and* $c_0 > c_0^*$

In this situation the domain conflict variation appeared in the i^{th} evidence instead of the $n+1^{\mathrm{th}}$. When we combine the available evidence we get a new basic probability assignment with

$$\forall \chi^* . m^*(e_q \notin (\vee \chi^*)) = \prod_{\chi_j \in \chi^*} m(e_q \notin \chi_j) \cdot \prod_{\chi_j \in (\chi - \chi^*)} [1 - m(e_q \notin \chi_j)]$$

but here $\chi^* \in 2^{\chi}$ where $\chi = \{\chi_1, \ldots, \chi_n\}$.

With no evidence from a $n+1^{\mathrm{th}}$ subset and domain conflict variation in the i^{th} evidence we have a slight change in the calculation of conflict,

$$k = m^*(e_q \notin (\vee \chi)) = \prod_{j=1}^{n} m(e_q \notin \chi_j) = \frac{c_0 - c_0^*}{1 - c_0} \cdot \prod_{\substack{j=1 \\ \neq i}}^{n} \frac{c_j^* - c_j}{1 - c_j}$$

and in the calculation of plausibility. For subsets except χ_i we get

$$\forall k \neq i.\,\mathrm{Bel}(e_q \in \chi_k) = 0,$$

$$\forall k \neq i.\,\mathrm{Pls}(e_q \in \chi_k) = 1 - \mathrm{Bel}(e_q \notin \chi_k) = 1 - \sum_{X \subseteq (e_q \notin \chi_k)} m(X)$$

$$= \frac{1 - m(e_q \notin \chi_k)}{1 - \prod_{j=1}^{n} m(e_q \notin \chi_j)} = \frac{1 - \dfrac{c_k^* - c_k}{1 - c_k}}{1 - \dfrac{c_0 - c_0^*}{1 - c_0^*} \cdot \prod_{\substack{j=1 \\ \neq i}}^{n} \dfrac{c_j^* - c_j}{1 - c_j}}$$

and for χ_i

$$\mathrm{Bel}(e_q \in \chi_i) = 0,$$

$$\mathrm{Pls}(e_q \in \chi_i) = 1 - \mathrm{Bel}(e_q \notin \chi_i) = 1 - \sum_{X \subseteq (e_q \notin \chi_i)} m(X)$$

$$= \frac{1 - m(e_q \notin \chi_i)}{1 - \prod_{j=1}^{n} m(e_q \notin \chi_j)} = \frac{1 - \dfrac{c_0^* - c_0}{1 - c_0}}{1 - \dfrac{c_0^* - c_0}{1 - c_0} \cdot \prod_{\substack{j=1 \\ \neq i}}^{n} \dfrac{c_j^* - c_j}{1 - c_j}}$$

3. When $e_q \in \chi_i$, $|\chi_i| = 1$ and $c_0 < c_0^*$

The increase in domain conflict when e_q was moved out from χ_i introduced a new type of evidence supporting that e_q belongs to χ_i. This changes the resulting bpa from the previous situations when we combine all evidences regarding e_q. Our new bpa is

$$\forall \chi^*.\, m^*(e_q \notin (\vee \chi^*))$$

$$= [1 - m(e_q \in \chi_i)] \cdot \prod_{\chi_j \in \chi^*} m(e_q \notin \chi_j) \cdot \prod_{\chi_j \in (\chi^{-i} - \chi^*)} [1 - m(e_q \notin \chi_j)]$$

and

$$m^*(e_q \notin (\vee \chi^{-i})) = m(e_q \in \chi_i) + [1 - m(e_q \in \chi_i)] \cdot \prod_{\chi_j \in \chi^{-i}} m(e_q \notin \chi_j)$$

where $\chi^* \in 2^{\chi^{-i}} - \{\chi^{-i}\}$, $\chi^{-i} = \chi - \{\chi_i\}$ and $\chi = \{\chi_1, \ldots, \chi_n\}$.

Since we did not have an evidence indicating that e_q did not belong to χ_i we will never have any support for the impossible statement that e_q does not belong anywhere in the new bpa. Thus, we will always get a zero conflict when combining these evidences;

$$k = m^{*}(e_q \notin (\ \vee \chi)) = 0.$$

When calculating belief and plausibility for any subset other than χ_i we get

$$\forall k \neq i.\,\mathrm{Bel}(e_q \in \chi_k) = 0$$

and

$$\begin{aligned}\forall k \neq i.\,\mathrm{Pls}(e_q \in \chi_k) &= 1 - \mathrm{Bel}(e_q \notin \chi_k) = 1 - \sum_{X \subseteq (e_q \notin \chi_k)} m(X) \\ &= 1 - [m(e_q \in \chi_i) + \{1 - m(e_q \in \chi_i)\} \cdot m(e_q \notin \chi_k)] \\ &= [1 - m(e_q \in \chi_i)] \cdot [1 - m(e_q \notin \chi_k)] \\ &= \left(1 - \frac{c_0}{c_0^*}\right) \cdot \left(1 - \frac{c_k^* - c_k}{1 - c_k}\right)\end{aligned}$$

and for χ_i:

$$\begin{aligned}\mathrm{Bel}(e_q \in \chi_i) &= m(e_q \in \chi_i) + [1 - m(e_q \in \chi_i)] \cdot \prod_{\chi_j \in \chi^{-i}} m(e_q \notin \chi_j) \\ &= \frac{c_0}{c_0^*} + \left(1 - \frac{c_0}{c_0^*}\right) \cdot \prod_{\substack{j=1 \\ \neq i}}^{n} \frac{c_j^* - c_j}{1 - c_j}.\end{aligned}$$

and

$$\mathrm{Pls}(e_q \in \chi_i) = 1 - \mathrm{Bel}(e_q \notin \chi_i) = 1 - \sum_{X \subseteq (e_q \notin \chi_i)} m(X) = 1.$$

because of the lack of evidence against that e_q belongs to the i^{th} subset.

B. The evidences specified

With plausibilities for all propositions that the evidence is referring to some particular subset we may now make a partial specification of each piece of evidence. That is, we will have an "evidential interval" of belief and plausibility for each possible subset. Since e_q belonged to χ_i as a result of the iterative partitioning of evidences, there was the least support against this and thus we will have the highest plausibility in favor of the proposition that e_q is referring to subset i. An evidence nonspecific with regard to which event it is referring to may then be specified from

evidence q:
proposition:
action part: $A_f, A_g, \ldots, A_h$
event part: $E_i, E_j, \ldots, E_k$
$m(A_f) = p_f$
$m(A_g) = p_g$
...
$m(A_h) = p_h$
$m(\Theta) = 1 - p_f - p_g - \ldots - p_h$

to

evidence q:
proposition:
action part: $A_f, A_g, \ldots, A_h$
event part: $[\mathrm{Bel}(e_q \in \chi_i), \mathrm{Pls}(e_q \in \chi_i)]\,/E_i$, $[0, \mathrm{Pls}(e_q \in \chi_j)]\,/E_j, \ldots,$
$[0, \mathrm{Pls}(e_q \in \chi_k)]\,/E_k$
$m(A_f) = p_f$
$m(A_g) = p_g$
...
$m(A_h) = p_h$
$m(\Theta) = 1 - p_f - p_g - \ldots - p_h$

IV. HANDLING THE FALSITY OF EVIDENCES

In Sec. III we received support k for the statement that an evidence e_q did not belong to any of the subsets. Since this is impossible the statement implies to a degree k that e_q is a false evidence. If an evidence is known to be false we would disregard the evidence completely, and when we have no indication as to the possible falsity of the evidences we would take no additional action.

We would then like to pay less and less regard to an evidence the higher the degree is that the evidence is false, pay no attention to the evidence when it is certainly false, and leave the evidence unchanged when there is no indication as to its falsity. This can be done by using the discounting operation introduced by Lowrance et al.[8] The discounting operation was introduced to handle the case when the source of an evidence is lacking in credibility. The credibility of the source, α, became also the credibility of the evidence. The situation was handled by discounting each supported proposition of the evidence other than Θ with the credibility α and by adding the discounted mass to Θ;

$$m^{\%}(A_j) = \begin{cases} \alpha \cdot m(A_j), & A_j \neq \Theta \\ 1 - \alpha + \alpha \cdot m(\Theta), & A_j = \Theta \end{cases}.$$

We will use the same discounting operation in this case when there is a direct indication for each separate piece of evidence regardless of which source produced it. We will view the support of the false statement that an evidence e_q does not belong to any subset and cannot be put in a subset by itself, i.e. the conflict in Dempster's rule when combining all evidences regarding e_q, as identical to one minus the credibility of the evidence;

$$\alpha \triangleq 1 - m^*(e_q \notin (\quad \vee \chi)) = 1 - k.$$

Thus, a piece of evidence is discounted in relation to its degree of falsity.

It is obvious that the credibility used to discount a piece of evidence depends on the evidence itself. This should be no problem since the credibility originates from an evidence at a "higher" level that depends on e_q but will never be combined with e_q. Instead, it is used to discount e_q. Obviously, any discounting directed towards individual evidences and not all evidences from a particular source will depend on the evidence itself.

We should note that we must not repartition the set of all evidences after the first discounting of all evidences in order to receive new credibilities and perform a second discounting. The two evidences from which the two credibilities are originating would not be independent. Thus, making a second discounting of an evidence would violate the independence assumption of Dempster's rule since a double discounting corresponds to combining the two nonindependent evidences concerning the falsity of e_q and discounting e_q with the credibility of e_q derived from the combination;

$$\alpha_{12} = 1 - k_{12} = 1 - [1 - (1 - k_1) \cdot (1 - k_2)] = (1 - k_1) \cdot (1 - k_2) = \alpha_1 \cdot \alpha_2$$

where α_1 and α_2 are the two credibilities of the first and second discounting and α_{12} is the credibility derived from the combination of both corresponding evidences, and k_1 and k_2 are the two degrees of falsity of the first and second partitioning and k_{12} is the degree of falsity in the combination of the two evidences.

In fact, we should never repartition evidences after discounting, regardless of whether we plan to perform a second discounting or not. The discounting operation not only puts the evidences in order for continuing reasoning processes regarding the different events, but it also smooths out the conflicting differences between evidences which is the very basis of the conflict minimizing process when the set of evidences are partitioned into subsets. Since the discounting smooths out the differences between evidences that do not belong to the same event, a repartitioning would only increase the risk that evidences referring to different events would be partitioned into the same subset. Thus, we should never repartition the set of evidences after discounting evidences for falsity.

The evidence we specified in Sec. III may now be discounted to its degree of credibility:

evidence q:
proposition:
action part: $A_f, A_g, \ldots, A_h$
event part: $[\mathrm{Bel}(e_q \in \chi_i), \mathrm{Pls}(e_q \in \chi_i)] / E_i,\ [0, \mathrm{Pls}(e_q \in \chi_j)] / E_j, \ldots, [0, \mathrm{Pls}(e_q \in \chi_k)] / E_k$

$$m(A_f) = \alpha \cdot p_f$$
$$m(A_g) = \alpha \cdot p_g$$
$$\ldots$$
$$m(A_h) = \alpha \cdot p_h$$
$$m(\Theta) = 1 - \alpha \cdot p_f - \alpha \cdot p_g - \ldots - \alpha \cdot p_h$$

where α is the degree of credibility and $1 - \alpha$ is the degree of falsity of e_q.

V. FINDING USABLE EVIDENCES

Obviously, the next question to put is: Will our now specified and discounted piece of evidence be of use in a subsequent reasoning process concerning a particular event? If this piece of evidence can only belong to one subset then it is also usable in a subsequent reasoning process for that subset. Whether this is the case or not will be determined by the now specified event part of the evidence. If the evidence will be useful in the reasoning process as well is another question. That depends only on the action part of the evidence.

If a piece of evidence can belong to more than one subset it will clearly be uncertain if it belongs to our subset in question if indeed it is possible at all. We must find a measure of this uncertainty—a credibility that the evidence belongs to the subset. Before using the evidence in the reasoning process concerning our subset, we would like to calculate the credibility that the evidence belongs to the subset in question and then discount the evidence by its credibility. Obviously, an evidence that cannot possible belong to a subset χ_i should be discounted entirely in the subsequent reasoning process for that subset, while an evidence which cannot possibly belong to any other subset χ_j and is without any support whatsoever against χ_i should not be discounted at all when used in the reasoning process for χ_i. Thus, the degree to which an evidence can belong to a subset and no other subset corresponds to the importance the evidence should be allowed to play in that subset.

In order to find the credibility of an evidence in the reasoning process for some subset we must measure the uncertainty in the newly specified event part of the evidence. Measures of uncertainty in a single piece of evidence are usually measures of entropy. An especially useful kind of such measure is the measure of average total uncertainty[9,10]. This is a measure of entropy that measures both scattering and nonspecificity of evidence:

$$H(m) = \sum_{A \in \Theta} m(A) \cdot \text{Log}_2(|A|) - \sum_{A \in \Theta} m(A) \cdot \text{Log}_2(m(A)).$$

However, the average total uncertainty in which event an evidence e_q might be referring to is not exactly our concern. This measure applied to the new basic probability assignment resulting from the fusion of all derived evidences regarding to which subset e_q can belong (Sec. III.A) would give us an indication of how usable the evidence is in total towards all subsets. What we want is a high plausibility for the most likely subset, i.e. little support against that the evidence belongs to the subset. This is equivalent with preferring a minimal entropy $H(m)$, but how the remainder of the support is scattered among the other focal elements is of little concern to us. Actually, if we are to express some preference regarding the remainder of the support we would choose some uniform scattering among the other focal elements, i.e preferring as low as possible a plausibility for the second most likely subset. This is not equivalent with preferring a minimal entropy. When it comes to the specificity of the support against different subsets, we prefer such a support to be specific when it concerns other subsets and most preferably gives these subsets a low plausibility, and to have some nonspecificity when it concerns the most preferable subset giving it a plausibility as high as possible. Thus, our overall preference is not consistent with a minimal entropy.

We might be able to find some entropy-like measure of entropy difference between two parts of an evidence that could be maximized. But rather than going this route we will make some simple observations (axioms):

- If the plausibility that e_q belongs to some subset is zero we should discount it entirely.

- If the plausibility of a subset is one and the plausibility of all other subsets are zero we should not discount e_q at all when used in this subset.

- If the plausibility of a subset is α, while the belief is zero, and the plausibility of all other subsets are zero we should discount e_q to a credibility of α.

- If the plausibility of n different subsets are all one and the plausibility of all other subsets are zero we should discount e_q to a credibility of $1/n$.

- The credibility of e_q when used in a subset is greater or equal to the belief of the subset.

A function that satisfy these observations is the plausibility of the subset weighted by the portion of the plausibility for all subsets that this subset has received and by the portion of the still uncommitted belief.

The credibility α_j of e_q when e_q is used in χ_j can then be calculated as

$$\alpha_j = [1 - \mathrm{Bel}(e_q \in \chi_i)] \cdot \frac{[\mathrm{Pls}(e_q \in \chi_j)]^2}{\sum_k \mathrm{Pls}(e_q \in \chi_k)}, j \neq i$$

$$\alpha_i = \mathrm{Bel}(e_q \in \chi_i) + [1 - \mathrm{Bel}(e_q \in \chi_i)] \cdot \frac{[\mathrm{Pls}(e_q \in \chi_i)]^2}{\sum_k \mathrm{Pls}(e_q \in \chi_k)}$$

Here, $\mathrm{Bel}(e_q \in \chi_i)$ is equal to zero except when $e_q \in \chi_i$, $|\chi_i| = 1$ and $c_0 < c_0^*$.

The discounting we make of e_q should not be confused with the discounting we made in Sec.IV. That discounting was made "on principle" due to the derived evidence proposing to some degree that e_q was false. The discounting we are making here however, is merely a technical necessity in order to be able to use the evidence when we as users force an absolute specificity upon the event part of the evidence by placing it in one of the subsets.

After discounting each evidence to its new credibility in a particular subset the subsequent reasoning process could begin. Note that a piece of evidence could be used in several different subsets with an appropriate discounting, i.e. for a particular subset every piece of evidence that belongs to the subset with a plausibility above zero could be used in the reasoning process within that subset. Also, an evidence whose original event part only indicated one possible event, say E_j, and which now has a plausibility of one for some subsets χ_i might still have a credibility below one for χ_i. This should come as no surprise since our evidence might not have been completely certain, i.e. leaving some mass on Θ. Since the mass on Θ supports any event, our evidence is not completely certain regarding which event it is referring to, giving us a credibility below one for χ_i. Also, even if it was completely certain, then χ_i, whose meaning is determined by all evidences it contains, might not for certain be representing E_j.

VI. FINDING THE EVENT OF A SUBSET

When we begin our subsequent reasoning process in each subset, it will naturally be of vital importance to know to which event the subset is referring. This information is obtainable when the evidences in the subset have been combined. After the combination, each focal element of the final bpa will in addition to supporting some proposition regarding an action also be referring to one or more events where the proposed action may have taken place. We could simply sum up the support in favor of each event, calculate the plausibility of it and then form our opinion regarding which event the subset is referring to based on this result. However, this may cause a problem. It would certainly be possible that more than one subset has one and the same event as its most likely event. This situation can be avoided if we bring the problem to the metalevel where we simultaneously reason about all subsets, i.e. which subsets are referring to which events. In this analysis we use our domain knowledge stating that no more than one subset may be referring to an event. From each subset we have an evidence indicating which events it might be referring to. This evidence is directly derivable from the final bpa resulting from the combination of all evidences in the subset. We simply remove the information about action from each focal element in the final bpa while leaving the information about event unchanged. This may leave us with two or more focal elements supporting the same event or disjunction of events. The support for these focal elements are summed up and the focal elements are represented only once. That is, we receive a new evidence at the metalevel originating from the subset that is not paying any attention to actions but paying the same attention to events as the final bpa resulting from the combination of all evidences within this subset. Thus, we have the following bpa of the evidence originating from χ_i:

$$\forall E . m_{\chi_i}((\vee E)/\chi_i) = \sum_{\text{Event part of } A \text{ is } E} m_{\chi_i}(A)$$

where $E \in 2^{\{E_j\}}$. Here, of course, $\vee \{E_j\}$ is Θ.

Combining all bpa's from all different subsets with the restriction that any intersection in the combination yielding $E_k/\chi_i \wedge E_k/\chi_j$ is false eliminates the possible problem of having an event simultaneously assigned to two or more different subsets. This method has a much higher chance to give a clearly preferable answer regarding which events is represented by which subsets, than that of only viewing the evidences within a subset when trying to determine its event.

VII. AN EXAMPLE

Let us return to the problem of two possible burglaries described in our first article.[2] We will now reexamine this problem in view of the results of Secs. II-VI. Finally we make a comparison between an overconfident approach of only partitioning the evidences by minimizing the metaconflict function before we begin the reasoning process separately in each subset, and the refined approach of discounting for falsity and uncertainty in affiliation proposed in this article.

A. A Refined Analysis of the Bakers' Shops Burglaries

In this example we had evidence weakly specified in the sense that it is uncertain to which possible burglary the propositions are referring. We will try to specify these evidences by studying cluster conflict variations when a piece of evidence is moved from its subset to another subset or put into a new subset by itself. The problem we were facing was described as follows:[2]

Assume that a baker's shop at One Baker Street has been burglarized, event 1. *Let there also be some indication that a baker's shop across the street, at Two Baker Street, might have been burglarized, although no burglary has been reported, event* 2. *An experienced investigator estimates that a burglary has taken place at Two Baker Street with a probability of* 0.4. *We have received the following evidences. A credible witness reports that "a brown-haired man who is not an employee at the baker's shop committed the burglary at One Baker Street," evidence* 1. *An anonymous witness, not being aware that there might be two burglaries, has reported "a brown-haired man who works at the baker's shop committed the burglary at Baker Street," evidence* 2. *Thirdly, a witness reports having seen "a suspicious-looking red-haired man in the baker's shop at Two Baker Street," evidence* 3. *Finally, we have a fourth witness, this witness, also anonymous and not being aware of the possibility of two burglaries, reporting that the burglar at the Baker Street baker's shop was a brown-haired man. That is, for example:*

evidence 1*:*
  proposition:
    action part: BO
    event part: E_1*:*
$m(BO) = 0.8$
$m(\Theta) = 0.2$

evidence 2*:*
  proposition:
    action part: BI
    event part: E_1, E_2
$m(BI) = 0.7$
$m(\Theta) = 0.3$

evidence 3*:*
  proposition:
    action part: R
    event part: E_2*:*
$m(R) = 0.6$
$m(\Theta) = 0.4$

evidence 4*:*
  proposition:
    action part: B
    event part: E_1, E_2
$m(B) = 0.5$
$m(\Theta) = 0.5$

domain probability distribution:

$$m(E_i) = \begin{cases} 0.6, i = 1 \\ 0.4, i = 2 \\ 0, i \neq 1, 2 \end{cases} .$$

All evidences where originally put into one subset, χ_1. By minimizing the metaconflict function it was found best to partition the evidences into two subsets. The minimum of the metaconflict function was found when evidences one and four were moved from χ_1 into χ_2 while evidences two and three remained in χ_1. From the event parts of the evidences we were able to conclude that χ_1 corresponded to event 2 and χ_2 corresponded to event 1.

Let us now study the cluster conflict variations. The conflict in χ_1 was $c_1 = 0.42$, in χ_2 it was $c_2 = 0$, with a domain conflict of $c_0 = 0.6$. If e_1 now in χ_2 is moved out from χ_2 the conflict will drop to zero, $c_2^* = 0$. If e_1 is then moved into χ_1 its conflict increases to $c_1^* = 0.788$, and if e_1 is put into a subset by itself, χ_3, we will have a domain conflict of one, $c_0^* = 1$.

Thus, by the formulas of Sec. II.C we get

$$m(e_1 \notin \chi_1) = \frac{c_1^* - c_1}{1 - c_1} = 0.634, \qquad m(e_1 \notin \chi_2) = \frac{c_2 - c_2^*}{1 - c_2^*} = 0$$

and

$$m(e_1 \notin \chi_3) = \frac{c_0^* - c_0}{1 - c_0} = 1.$$

From these evidences we will calculate the plausibility for each subset that e_1 belongs to the subset:

$$\mathrm{Pls}(e_1 \in \chi_1) = \frac{1 - m(e_1 \notin \chi_1)}{1 - m(e_1 \notin \chi_1) \cdot m(e_1 \notin \chi_2) \cdot m(e_1 \notin \chi_3)} = 1 - 0.634 = 0.366,$$

$$\mathrm{Pls}(e_1 \in \chi_2) = \frac{1 - m(e_1 \notin \chi_2)}{1 - m(e_1 \notin \chi_1) \cdot m(e_1 \notin \chi_2) \cdot m(e_1 \notin \chi_3)} = 1,$$

$$\mathrm{Pls}(e_1 \in \chi_3) = \frac{1 - m(e_1 \notin \chi_3)}{1 - m(e_1 \notin \chi_1) \cdot m(e_1 \notin \chi_2) \cdot m(e_1 \notin \chi_3)} = 0.$$

We do the same for the other three evidences:

$$m(e_2 \notin \chi_i) = \begin{cases} 0.42, & i = 1 \\ 0.56, & i = 2 \\ 1, & i = 3 \end{cases}, \qquad m(e_3 \notin \chi_i) = \begin{cases} 0.42, & i = 1 \\ 0.54, & i = 2 \\ 1, & i = 3 \end{cases},$$

$$m(e_4 \notin \chi_i) = \begin{cases} 0.155, & i = 1 \\ 0, & i = 2 \\ 1, & i = 3 \end{cases}$$

and calculate the plausibilities

$$\mathrm{Pls}(e_2 \in \chi_i) = \begin{cases} 0.758, & i = 1 \\ 0.575, & i = 2 \\ 0, & i = 3 \end{cases}, \qquad \mathrm{Pls}(e_3 \in \chi_i) = \begin{cases} 0.750, & i = 1 \\ 0.595, & i = 2 \\ 0, & i = 3 \end{cases},$$

$$\mathrm{Pls}(e_4 \in \chi_i) = \begin{cases} 0.845, & i = 1 \\ 1, & i = 2 \\ 0, & i = 3 \end{cases} .$$

Thus, the four evidences are specified as:

evidence 1:
  proposition:
    action part: BO
    event part:
      $[0, 0.366] / \chi_1$, $[0, 1] / \chi_2$
$m(BO) = 0.8$
$m(\Theta) = 0.2$

evidence 2:
  proposition:
    action part: BI
    event part:
      $[0, 0.758] / \chi_1$, $[0, 0.575] / \chi_2$
$m(BI) = 0.7$
$m(\Theta) = 0.3$

evidence 3:
  proposition:
    action part: R
    event part:
      $[0, 0.750] / \chi_1$, $[0, 0.595] / \chi_2$
$m(R) = 0.6$
$m(\Theta) = 0.4$

evidence 4:
  proposition:
    action part: B
    event part:
      $[0, 0.845] / \chi_1$, $[0, 1] / \chi_2$
$m(B) = 0.5$
$m(\Theta) = 0.5$

Thus, it seems pretty certain that e_1 belongs to χ_2 while the other three evidences are more uncertain in which subset they belong to, i.e. more nonspecific in which event they are referring to. Especially e_4 is not specific. It could almost belong to either subset.

When we combined the evidences regarding where a particular evidence might belong, we received a conflict for e_2 and e_3 but not for e_1 and e_4. Thus, there is no indication that e_1 and e_4 might be false. For the second and third evidence we got a conflict of 0.2352 and 0.2268 respectively. This is their degrees of falsity. We should then discount e_2 and e_3 to their respective degrees of credibility, i.e. 0.7648 and 0.7732:

evidence 1:
  proposition:
    action part: BO
    event part:
      $[0, 0.366] / \chi_1$, $[0, 1] / \chi_2$
$m(BO) = 0.8$
$m(\Theta) = 0.2$

evidence 2:
  proposition:
    action part: BI
    event part:
      $[0, 0.758] / \chi_1$, $[0, 0.575] / \chi_2$
$m(BI) = 0.5354$
$m(\Theta) = 0.4646$

evidence 3:
proposition:
action part: R
event part:
$[0, 0.750]/\chi_1$, $[0, 0.595]/\chi_2$
$m(R) = 0.4639$
$m(\Theta) = 0.5361$

evidence 4:
proposition:
action part: B
event part:
$[0, 0.845]/\chi_1$, $[0, 1]/\chi_2$
$m(B) = 0.5$
$m(\Theta) = 0.5$

The discounting of e_2 and e_3 due to their fairly high degree of falsity will reduce the impact of these two evidences in a subsequent reasoning process regarding the two different events.

Before we finally start the reasoning process in χ_1 and χ_2 we should once again discount the evidences. This time we make an individual discounting for each subset and evidence according to how credible it is that the evidence belongs to the subset. The credibility that e_1 belongs to χ_1 is

$$\alpha_1 = \frac{(\text{Pls}(e_1 \in \chi_1))^2}{\sum_{j=1}^{2} \text{Pls}(e_1 \in \chi_j)} = \frac{0.366^2}{0.366 + 1} = 0.0981$$

and that e_1 belongs to χ_2

$$\alpha_2 = \frac{(\text{Pls}(e_1 \in \chi_2))^2}{\sum_{j=1}^{2} \text{Pls}(e_1 \in \chi_j)} = \frac{1}{0.366 + 1} = 0.7321.$$

For the other three evidences we get: e_2: $\alpha_1 = 0.4310$, $\alpha_2 = 0.2480$, e_3: $\alpha_1 = 0.4182$, $\alpha_2 = 0.2632$, and for e_4: $\alpha_1 = 0.3870$, $\alpha_2 = 0.5420$. Discounting the four evidences to their credibility of belonging to χ_1 and χ_2, respectively, yields:

evidence 1:
proposition:
action part: BO
event part:
$[0, 0.366]/\chi_1$, $[0, 1]/\chi_2$
$m(BO) = 0.0784/\chi_1, 0.5856/\chi_2$
$m(\Theta) = 0.9216/\chi_1, 0.4144/\chi_2$

evidence 2:
proposition:
action part: BI
event part:
$[0, 0.758]/\chi_1$, $[0, 0.575]/\chi_2$
$m(BI) = 0.2308/\chi_1, 0.1328/\chi_2$
$m(\Theta) = 0.7692/\chi_1, 0.8672/\chi_2$

evidence 3:
proposition:
action part: R
event part:
$[0, 0.750]/\chi_1$, $[0, 0.595]/\chi_2$
$m(R) = 0.1940/\chi_1, 0.1221/\chi_2$
$m(\Theta) = 0.8060/\chi_1, 0.8779/\chi_2$

evidence 4:
proposition:
action part: B
event part:
$[0, 0.9216]/\chi_1$, $[0, 0.4144]/\chi_2$
$m(B) = 0.1935/\chi_1, 0.2710/\chi_2$
$m(\Theta) = 0.8065/\chi_1, 0.7290/\chi_2$

Combining these four evidences with Dempster's rule results in the following final basic probability assignment:

$$m^*_{1 \oplus 2 \oplus 3 \oplus 4}(BO \wedge E_1)$$

$$= \frac{1}{1-k} \cdot m_1(BO \wedge E_1) \cdot [1 - m_2(BI \wedge (E_1 \vee E_2))] \cdot [1 - m_3(R \wedge E_2)]$$

$$= 0.0539/\chi_1, 0.5298/\chi_2,$$

$$m^*_{1 \oplus 2 \oplus 3 \oplus 4}(BI \wedge (E_1 \vee E_2))$$

$$= \frac{1}{1-k} \cdot [1 - m_1(BO \wedge E_1)] \cdot m_2(BI \wedge (E_1 \vee E_2)) \cdot [1 - m_3(R \wedge E_2)]$$

$$= 0.1900/\chi_1, 0.0574/\chi_2,$$

$$m^*_{1 \oplus 2 \oplus 3 \oplus 4}(B \wedge (E_1 \vee E_2))$$

$$= \frac{1}{1-k} \cdot [1 - m_1(BO \wedge E_1)] \cdot [1 - m_2(BI \wedge (E_1 \vee E_2))] \cdot [1 - m_3(R \wedge E_2)]$$

$$\cdot m_4(B \wedge (E_1 \vee E_2)) = 0.1225/\chi_1, 0.1016/\chi_2,$$

$$m^*_{1 \oplus 2 \oplus 3 \oplus 4}(R \wedge E_2)$$

$$= \frac{1}{1-k} \cdot [1 - m_1(BO \wedge E_1)] \cdot [1 - m_2(BI \wedge (E_1 \vee E_2))] \cdot m_3(R \wedge E_2)$$

$$\cdot [1 - m_4(B \wedge (E_1 \vee E_2))] = 0.1229/\chi_1, 0.0380/\chi_2,$$

$$m^*_{1 \oplus 2 \oplus 3 \oplus 4}(\Theta)$$

$$= \frac{1}{1-k} \cdot [1 - m_1(BO \wedge E_1)] \cdot [1 - m_2(BI \wedge (E_1 \vee E_2))] \cdot [1 - m_3(R \wedge E_2)]$$

$$\cdot [1 - m_4(B \wedge (E_1 \vee E_2))] = 0.5107/\chi_1, 0.2732/\chi_2.$$

where k is the conflict in Dempster's rule;

$$k = m_3(R \wedge E_2) \cdot \Bigg(1 - [1 - m_1(BO \wedge E_1)]$$

$$\cdot\ [1 - m_2(BI \wedge (E_1 \vee E_2))] \cdot [1 - m_4(B \wedge (E_1 \vee E_2))]\Bigg)$$

$$+\ [1 - m_3(R \wedge E_2)] \cdot m_1(BO \wedge E_1) \cdot m_2(BI \wedge (E_1 \vee E_2)) = 0.0977/\chi_1, 0.1584/\chi_2.$$

Finally, this gives us the following evidential intervals:

$$\begin{aligned}
[\mathrm{Bel}(BO), \mathrm{Pls}(BO)] &= [0.0539, 0.6871]\,/\chi_1,\ [0.5298, 0.9046]\,/\chi_2,\\
[\mathrm{Bel}(BI), \mathrm{Pls}(BI)] &= [0.1900, 0.8232]\,/\chi_1,\ [0.0574, 0.4322]\,/\chi_2,\\
[\mathrm{Bel}(B), \mathrm{Pls}(B)] &= [0.3664, 0.8771]\,/\chi_1,\ [0.6888, 0.9620]\,/\chi_2,\\
[\mathrm{Bel}(R), \mathrm{Pls}(R)] &= [0.1229, 0.6336]\,/\chi_1,\ [0.0380, 0.3112]\,/\chi_2,\\
[\mathrm{Bel}(I), \mathrm{Pls}(I)] &= [0.1900, 0.9461]\,/\chi_1,\ [0.0574, 0.4702]\,/\chi_2,\\
[\mathrm{Bel}(O), \mathrm{Pls}(O)] &= [0.0539, 0.8100]\,/\chi_1,\ [0.5298, 0.9426]\,/\chi_2,\\
[\mathrm{Bel}(E_1), \mathrm{Pls}(E_1)] &= [0.0539, 0.8771]\,/\chi_1,\ [0.5298, 0.9620]\,/\chi_2,\\
[\mathrm{Bel}(E_2), \mathrm{Pls}(E_2)] &= [0.1229, 0.9461]\,/\chi_1,\ [0.0380, 0.4702]\,/\chi_2.
\end{aligned}$$

From the intervals regarding which event the subsets are referring to it is somewhat uncertain whether χ_1 is referring to E_1 or E_2. However, χ_2 clearly refers to E_1.

Let us bring the problem to the metalevel together with our domain knowledge that the two subsets must be referring to different events. We create two new but very similar basic probability assignments as follows:

$$\begin{aligned}
m_{\chi_1}(E_1/\chi_1) &= 0.0539,\\
m_{\chi_1}(E_2/\chi_1) &= 0.1229,\\
m_{\chi_1}(\Theta) = 1 - m_{\chi_1}(E_1/\chi_1) - m_{\chi_1}(E_2/\chi_1) &= 0.8232
\end{aligned}$$

and

$$\begin{aligned}
m_{\chi_2}(E_1/\chi_2) &= 0.5298,\\
m_{\chi_2}(E_2/\chi_2) &= 0.0380,\\
m_{\chi_2}(\Theta) = 1 - m_{\chi_2}(E_1/\chi_2) - m_{\chi_2}(E_2/\chi_1) &= 0.4322.
\end{aligned}$$

Combining these on the metalevel yields

$$m_{\chi_1 \oplus \chi_2}(E_1/\chi_1 \wedge E_2/\chi_2) = 0.0699,$$
$$m_{\chi_1 \oplus \chi_2}(E_2/\chi_1 \wedge E_1/\chi_2) = 0.6840,$$
$$m_{\chi_1 \oplus \chi_2}(\Theta) = 0.2462$$

with evidential intervals

$$[\mathrm{Bel}(E_1/\chi_1 \wedge E_2/\chi_2), \mathrm{Pls}(E_1/\chi_1 \wedge E_2/\chi_2)] = [0.0699, 0.3160]$$

and

$$[\mathrm{Bel}(E_2/\chi_1 \wedge E_1/\chi_2), \mathrm{Pls}(E_2/\chi_1 \wedge E_1/\chi_2)] = [0.6840, 0.9301]$$

This makes it perfectly clear that χ_1 refers to E_2 while χ_2 refers to E_1.

We see in conclusion that at χ_1, i.e. event 2, there is some support for the burglar being brown-haired although it is certainly plausible, although less likely, he was actually red-haired. We have an even slighter indication that this might be an inside job but it is also possible that the burglar was an outsider. In general the evidence regarding event 1 is pretty inconclusive. However, the picture is much clearer at χ_2, i.e. event 1. It is quite likely that the burglar at event 1 was a brown-haired outsider.

B. A Comparison Between an Overconfident and a Refined Analysis of the Bakers' Shops Burglaries

When we partitioned the four evidences, e_2 and e_3 ended up in χ_1 while e_1 and e_4 ended up in χ_2. This was the partitioning that minimized the metaconflict and thus the most probable partition. However, it said nothing about the probability for some evidence that it belonged to the subset where it was placed, and nothing about how much more probable this subset was to other subsets. It only said that this was the most probable subset of all. Thus, an evidence might end up in some subset that was only marginally better than some other. This somewhat overconfident approach might then falsely indicate a certainty in the subsequent reasoning process within each subset that does not really exist. This false certainty is due to the restriction of not, to any degree, using evidences that ended up in other subsets by the partitioning. In contrast, the refined approach uses all evidences that could possibly belong to a subset in the reasoning process for that subset, although they are discounted to the credibility that they belong to the subset. This approach eliminates the problem of false certainty imposed by the partitioning as seen in the following comparison of the two approaches applied to the bakers' shops burglary problem.

As a comparison between the two approaches there is not much to say about the conclusions drawn in χ_2. Whatever was concluded earlier is also concluded in the refined approach. That is, our burglar is a brown-haired outsider. The only real difference seems to be a somewhat higher plausibility for unlikely read-haired and insider alternatives together with a lower support for the preferred brown-haired and outsider alternatives due mainly to the possibility that e_2 or e_3 placed in χ_1 by the partitioning might belong to the subset.

As before, the situation is not so clear at χ_1. In general we find that evidential intervals have opened up in the refined approach. This is due to the discounting of evidence. In the refined approach we see an especially large drop in support for alternatives supported by evidences that belonged to χ_1 in the overconfident approach, brown-haired insider and red-haired, and also a large increase in plausibility for alternatives supported by evidences that belonged to χ_2, brown-haired and brown-haired outsider. This is due to the possibility that the evidences that belonged to χ_2 in the overconfident approach actually has a possibility of belonging instead to χ_1, and vice versa. If we consider the three alternatives brown-haired insider, brown-haired and insider in χ_1 they all had a support of 0.483 and a plausibility of 0.69, 0.69 and 1, respectively, in the overconfident approach. In the refined approach there is only a small drop in support for brown-haired to 0.36 but a much larger drop to 0.19 for insider and brown-haired insider. This is due to the possibility that e_1 supporting brown-haired outsider belongs to χ_1. This might not be very plausible but if it was the case it would have a large impact since e_1 is strongly supportive of brown-haired outsider. Thus, in the overconfident approach we might have falsely concluded that the burglar was a brown-haired insider while it actually, as shown in the refined approach, is much more of an open question whether the probably brown-haired burglar was an insider or not.

VIII. CONCLUSIONS

In this article we have extended the methodology to partition nonspecific evidence developed in our previous article[2] to a methodology for specifying nonspecific evidence. This is in itself clearly an important extension in analysis, considering that an evidence will now in a subsequent reasoning process be handled similarly by different subsets if these are approximately equally plausible, whereas before the most plausible subset would take the evidence as certainly belonging to the subset while the other subsets would never consider the evidence in their reasoning processes. In addition, two facts will facilitate the reasoning process. Firstly, the specification process in the extended methodology will besides specifying the evidences also give a degree of falsity and a degree of credibility in affiliation for each piece of evidence. Secondly, the methodology can iteratively receive its evidences piece by piece. Together, these facts indicates that it should be possible to develop methods for disregarding immediately upon receipt false evidence as well as methods for focusing attention upon useful evidence based on their maximum degree of credibility.

I would like to thank Stefan Arnborg, Ulla Bergsten and Per Svensson for their helpful comments regarding this article.

References

1. J.D. Lowrance and T.D. Garvey, "Evidential Reasoning: An Implementation for Multisensor Integration," Technical Note 307, Artificial Intelligence Center, Computer Science and Technology Division, SRI International, Menlo Park, CA, 1983.
2. J. Schubert, "On Nonspecific Evidence," *Int. J. Intell. Syst.* **8**, 711-725 (1993).
3. A.P. Dempster, "A generalization of Bayesian inference," *J. R. Stat. Soc. Ser. B* **30**, 205-247 (1968).
4. G. Shafer, "A Mathematical Theory of Evidence" (Princeton University, Princeton, 1976).
5. G. Shafer, "Perspectives on the Theory and Practice of Belief Functions," *Int. J. Approx. Reason.* **4**, 323-362, (1990).
6. G. Shafer, "Rejoinders to comments on 'Perspectives on the Theory and Practice of Belief Functions'," *Int. J. Approx. Reason.* **6**, 445-480, (1992).
7. U. Bergsten, and J. Schubert, "Dempster's Rule for Evidence Ordered in a Complete Directed Acyclic Graph," *Int. J. Approx. Reason.* **9**, 37-73 (1993).
8. J.D. Lowrance, T.D. Garvey, and T.M. Strat, "A framework for evidential-reasoning systems," in *Proc. 5th Nat. Conf. AI,* Philadelphia, PA, 1986, pp. 896-901.
9. N.R. Pal, J.C. Bezdek, and R. Hemasinha, "Uncertainty Measures for Evidential Reasoning I: A Review," *Int. J. Approx. Reason.* **7**, 165-183 (1992).
10. N.R. Pal, J.C. Bezdek, and R. Hemasinha, "Uncertainty Measures for Evidential Reasoning II: A New Measure of Total Uncertainty," *Int. J. Approx. Reason.* **8**, 1-16 (1993).

Finding a Posterior Domain Probability Distribution by Specifying Nonspecific Evidence

Manuscript

III

FINDING A POSTERIOR DOMAIN PROBABILITY DISTRIBUTION BY SPECIFYING NONSPECIFIC EVIDENCE

JOHAN SCHUBERT
Division of Information System Technology,
Department of Command and Control Warfare Technology,
National Defence Research Establishment,
S-172 90 Stockholm, SWEDEN

This article is an extension of the results of two earlier articles. In [J. Schubert, "On nonspecific evidence", *Int. J. Intell. Syst.* **8** (1993) 711-725] we established within Dempster-Shafer theory a criterion function called the metaconflict function. With this criterion we can partition into subsets a set of evidences with propositions that are weakly specified in the sense that it may be uncertain to which event a proposition is referring. In a second article [J. Schubert, "Specifying nonspecific evidence", FOA Report C 20975-2.7, National Defence Research Establishment, Stockholm, 1994] we not only found the most plausible subset for each piece of evidence, we also found the plausibility for every subset that the evidence belongs to the subset. In this article we aim to find a posterior probability distribution regarding the number of subsets. We use the idea that each evidence in a subset supports the existence of that subset to the degree that the evidence supports anything at all. From this we can derive a bpa that is concerned with the question of how many subsets we have. That bpa can then be combined with a given prior domain probability distribution in order to obtain the sought-after posterior domain distribution.

Keywords: belief functions, Dempster-Shafer theory, evidential reasoning, evidence correlation, cluster analysis, posterior distribution.

1. Introduction

In two earlier articles [1, 2] we derived methods, within the framework of Dempster-Shafer theory [3-6], to handle evidences that are weakly specified in the sense that it may not be certain to which of several possible events a proposition is referring. When reasoning with such evidences we must avoid combining evidences by mistake that refer to different events.

The methodology developed in these two articles was intended for a multiple-target tracking algorithm in an anti-submarine intelligence analysis system [1, 7]. In this application a sparse flow of intelligence reports arrives at the analysis system. These reports may originate from several different unconnected sensor systems. The reports carry a proposition about the occurrence of a submarine at a specified time and place, a probability of the truthfulness of the report and may contain additional information such as velocity, direction and type of submarine.

When there are several submarines we want to separate the intelligence reports into subsets according to which submarine they are referring to. We will then analyze the reports for each submarine separately. However, the intelligence reports are never labeled as to which submarine they are referring to. Thus, it is not possible to directly differentiate between two different submarines using two intelligence reports.

Instead we will use the conflict between the propositions of two intelligence reports as a probability that the two reports are referring to different submarines. This probability is the basis for separating intelligence reports into subsets.

The cause of the conflict can be non-firing sensors placed between the positions of the two reports, the required velocity to travel between the positions of the two reports at their respective times in relation to the assumed velocity of the submarines, etc.

In the first of these two articles we partitioned the set of evidences into subsets, where each subset was representing a separate event. These subsets should then be handled separately by subsequent reasoning processes. This methodology was able to find the optimal partitioning of evidence among subsets as well as the optimal estimate of the number of subsets when our own domain knowledge regarding the actual number of subsets was uncertain.

In the second article we found support regarding each piece of evidence and every subset that the evidence does not belong to the subset. This support is used to specify each piece of evidence, in the sense that we find to which events the proposition of the evidence might be referring, by calculating the belief and plausibility for each subset that the evidence belong to the subset. During this evidence specifying process we receive indications that some evidence might be false. Also, it became apparent that some evidences might not be so useful since they could belong to several different subsets. These evidences were discounted and were not allowed to strongly influence a subsequent reasoning process within a subset.

In this article we extend the work described in [1, 2] and aim to find a posterior probability distribution regarding the number of subsets by combining a given prior distribution with evidence regarding the number of subsets received from the evidence specifying process. We use the idea that each evidence in a subset supports the existence of that subset to the degree that that evidence supports anything at all. The evidences in each subset are combined and the resulting evidence is the total support for the subset. However, for every original piece of evidence in the subset we have a second evidence, derived in [2], with a proposition that supports that this evidence does not belong to the subset. If we have such support for every single piece of evidence in some subset, then this is also support that the subset itself is false. Thus, in this case, we will discount the evidence that supports the existence of the subset. Such discounted evidences that support the existence of different subsets, one from each subset, are then combined.

From the resulting basic probability assignment (bpa) of that combination we can create a new bpa by exchanging each and every proposition. A proposition in the new bpa is then a statement about the existence of a minimal number of subsets where this number is the length of a conjunction of terms of the corresponding proposition in the previous bpa. Thus, where the previous bpa is concerned with the question of which subsets have support, the new bpa is concerned with the question of how many subsets are supported. The new bpa gives us some opinion, based only on the evidence specifying process, about the probability of different numbers of subsets.

In order to obtain the sought-after posterior domain probability distribution we combine the bpa from the evidence specifying process with the given prior distribution from the problem specification.

In Section 2 of this article we give a summary of the two previous articles [1, 2]. We investigate in Section 3 what domain relevant conclusions can be drawn from the evidence specifying process and then derive the posterior distribution. Finally, in Section 4, we give a detailed example.

2. Summary of articles [1, 2]

In this summary we will focus on results of the previous two articles that we need to derive a posterior domain probability distribution regarding the number of events. It will be derived by a combination of a given prior probability distribution and a bpa resulting from an evidence specification process [2] where we study the changes in conflict when we move an evidence from one subset to another.

However, first we will learn how to separate evidences based on their conflicts [1]. Since our evidences are weakly specified with respect to which events they are referring, it is impossible to directly separate evidences based only on their proposition. The conflict in Dempster's rule measures the lack of compatibility between evidences. Since evidences referring to different events tend to be more incompatible than evidences referring to the same event, it is an obvious choice as a distance measure between evidences in a cluster algorithm. The idea of using the conflict in Dempster's rule as distance measure between evidences was first suggested by Lowrance and Garvey [8].

2.1. *On nonspecific evidence [1]*

In [1] we established a criterion function of overall conflict called the metaconflict function. With this criterion we can partition evidences with weakly specified propositions into subsets, each subset representing a separate event. These events should be handled independently.

To make a separation of evidences possible, every proposition's action part must be supplemented with an event part describing to which event the proposition is referring. If the proposition is written as a conjunction of literals or disjunctions, then one literal or disjunction concerns which event the proposition is referring to. This is the event part. The remainder of the proposition is called the action part. An example from our earlier article illustrates the terminology:

Let us consider the burglaries of two bakers' shops at One and Two Baker Street, event 1 (E_1) and event 2 (E_2), i.e., the number of events is known to be two. One witness hands over an evidence, specific with respect to event, with the proposition: "The burglar at One Baker Street," event part: E_1, "was probably brown haired (B)," action part: B. A second anonymous witness hands over a nonspecific evidence with the proposition: "The burglar at Baker Street," event part: E_1, E_2, "might have been red haired (R)," action part: R. That is, for example:

evidence 1:
proposition:
action part: B
event part: E_1:
$m(B) = 0.8$
$m(\Theta) = 0.2$

evidence 2:
proposition:
action part: R
event part: E_1,E_2
$m(R) = 0.4$
$m(\Theta) = 0.6$

2.1.1. *Separating nonspecific evidence*

We will have a conflict between two pieces of evidence in the same subset in two different situations. First, we have a conflict if the proposition action parts are conflicting regardless of the proposition event parts since they are presumed to be referring to the same event. Secondly, if the proposition event parts are conflicting then, regardless of the proposition action parts, we have a conflict with the presumption that they are referring to the same event.

The metaconflict used to partition the set of evidences is derived as the plausibility that the partitioning is correct when the conflict in each subset is viewed as a metalevel evidence against the partitioning of the set of evidences, χ, into the subsets, χ_i. We have a simple frame of discernment on the metalevel $\Theta = \{\text{AdP}, \neg\text{AdP}\}$, where AdP is short for "adequate partition", and a bpa from each subset χ_i assigning support to a proposition against the partitioning:

$$m_{\chi_i}(\neg\text{AdP}) \triangleq \text{Conf}(\{e_j | e_j \in \chi_i\}),$$

$$m_{\chi_i}(\Theta) \triangleq 1 - \text{Conf}(\{e_j | e_j \in \chi_i\})$$

where e_j is the jth evidence and $\{e_j | e_j \in \chi_i\}$ is the set of evidences belonging to subset χ_i and $\text{Conf}(\cdot)$ is the conflict, k, in Dempster's rule. Also, we have a bpa concerning the domain resulting from a probability distribution about the number of subsets, E, conflicting with the actual current number of subsets, $\#\chi$. This bpa also assigns support to a proposition against the partitioning:

$$m_D(\neg \text{AdP}) \triangleq \text{Conf}(\{E, \#\chi\}),$$

$$m_D(\Theta) \triangleq 1 - \text{Conf}(\{E, \#\chi\}).$$

The combination of these by Dempster's rule give us the following plausibility of the partitioning:

$$\text{Pls}(\text{AdP}) = (1 - m_D(\neg \text{AdP})) \cdot \prod_{i=1}^{r} (1 - m_{\chi_i}(\neg \text{AdP})).$$

The difference, one minus the plausibility of a partitioning, will be called the metaconflict of the partitioning.

2.1.2. *Metaconflict as a criterion function*

The metaconflict function can then be defined as:

DEFINITION. *Let the* metaconflict function,

$$Mcf(r, e_1, e_2, \ldots, e_n) \triangleq 1 - (1 - c_0) \cdot \prod_{i=1}^{r} (1 - c_i),$$

be the conflict against a partitioning of n evidences of the set χ *into r disjoint subsets* χ_i *where*

$$c_0 = \sum_{i \neq r} m(E_i)$$

is the conflict between r subsets and propositions about possible different number of subsets, E_i *the proposition that there are i subsets,* $m(E_i)$ *the support for it and*

$$c_i = \sum_{\substack{I \\ \cap I = \varnothing}} \prod_{e_j^k \in I} m_j(e_j^k)$$

is the conflict in subset i, where $\cap I$ *is the intersection of all elements in I,* $I = \{e_j^k | e_j \in \chi_i\}$ *is a set of one focal element from the support function of each evidence* e_j *in* χ_i *and* e_j^k *is the* kth *focal element of evidence* e_j.

Thus, $|I| = |\chi_i|$ and

$$|\{I\}| = \prod_{e_j \in \chi_i} |e_j|$$

where $|e_j|$ is the number of focal elements of e_j.

Two theorems are derived to be used in the separation of the set of evidences into subsets by an iterative minimization of the metaconflict function. By using these theorems we are able to reason about the optimal estimate of the number of events, when the actual number of events may be uncertain, as well as the optimal partition of nonspecific evidence for any fixed number of events. These two theorems will also be useful in a process for specifying evidences by observing changes in the metaconflict when moving a single piece of evidences between different subsets.

THEOREM 1. *For all j with* $j < r$, *if* $m(E_j) < m(E_r)$ *then* $\min Mcf(r, e_1, e_2, \ldots, e_n) < \min Mcf(j, e_1, e_2, \ldots, e_n)$.

This theorem states that an optimal partitioning for r subsets is always better than the other solutions with fewer than r subsets if the basic probability assignment for r subsets is greater than the basic probability assignment for the fewer subsets.

THEOREM 2. *For all j, if* $\min Mcf(r, e_1, e_2, \ldots, e_n) < \sum_{i \neq j} m(E_i)$ *then* $\min Mcf(r, e_1, e_2, \ldots, e_n) < \min Mcf(j, e_1, e_2, \ldots, e_n)$.

Theorem 2 states that an optimal partitioning for some number of subsets is always better than other solutions for any other number of subsets when the domain part of the metaconflict function is greater than the total metaconflict of the present partitioning.

2.2. *Specifying nonspecific evidence [2]*

2.2.1. *Evidences about evidence*

A conflict in a subset χ_i is interpreted as an evidence that there is at least one piece of evidence that does not belong to the subset;

$$m_{\chi_i}(\exists j . e_j \notin \chi_i) = c_i.$$

If an evidence e_q in χ_i is taken out from the subset the conflict c_i in χ_i decreases to c_i^*. This decrease $c_i - c_i^*$ was interpreted as an evidence indicating that e_q does not belong to χ_i, $m_{\Delta\chi_i}(e_q \notin \chi_i)$, and the remaining conflict c_i^* is an other evidence indicating that there is at least one other evidence e_j, $j \neq q$, that does not belong to χ_i - $\{e_q\}$,

$$m_{\chi_i - \{e_q\}}(\exists j \neq q . e_j \notin (\chi_i - \{e_q\})) = c_i^*.$$

The unknown bpa, $m_{\Delta\chi_i}(e_q \notin \chi_i)$, was derived by stating that the belief that there is at least one piece of evidence that does not belong to χ_i should be equal, no matter whether that belief is based on the original evidence $m_{\chi_i}(\exists j.\, e_j \notin \chi_i)$, before e_q is taken out from χ_i, or on a combination of the other two evidences $m_{\Delta\chi_i}(e_q \notin \chi_i)$ and $m_{\chi_i - \{e_q\}}(\exists j \neq q.\, e_j \notin (\chi_i - \{e_q\}))$, after e_q is taken out from χ_i, i.e.

$$\mathrm{Bel}_{\chi_i}(\exists j.\, e_j \notin \chi_i) = \mathrm{Bel}_{\Delta\chi_i \oplus (\chi_i - \{e_q\})}(\exists j.\, e_j \notin \chi_i).$$

where

$$\mathrm{Bel}_{\chi_i}(\exists j.\, e_j \notin \chi_i) = c_i$$

and

$$\mathrm{Bel}_{\Delta\chi_i \oplus (\chi_i - \{e_q\})}(\exists j.\, e_j \notin \chi_i) = c_i^* + m_{\Delta\chi_i}(e_q \notin \chi_i) \cdot [1 - c_i^*].$$

Thus, we derived an evidence that e_q does not belong to χ_i from the variations in cluster conflict when e_q was taken out from χ_i:

$$m_{\Delta\chi_i}(e_q \notin \chi_i) = \frac{c_i - c_i^*}{1 - c_i^*}.$$

If e_q after it is taken out from χ_i is brought into another subset χ_k, its conflict will increase from c_k to c_k^*. The increase in conflict when e_q is brought into χ_k is interpreted as if there exists some evidence indicating that e_q does not belong to $\chi_k + \{e_q\}$, i.e.

$$\forall k \neq i.\, m_{\Delta\chi_k}(e_q \notin (\chi_k + \{e_q\})) = \frac{c_k^* - c_k}{1 - c_k}.$$

When we take out an evidence e_q from subset χ_i and move it to some other subset we might have a changes in domain conflict. The domain conflict was interpreted as an evidence that there exists at least one piece of evidence that does not belong to any of the n first subsets, $n = |\chi|$, or if that particular evidence was in a subset by itself, as an evidence that it belongs to one of the other n-1 subsets. This indicate that the number of subsets is incorrect.

When $|\chi_i| > 1$ we may not only put an evidence e_q that we have taken out from χ_i into another already existing subset, we may also put e_q into a new subset χ_{n+1} by itself. There is no change in the domain conflict when we take out e_q from χ_i since $|\chi_i| > 1$, thus we may interpret the domain conflict as

$$m_\chi(\exists j \neq q \forall k \neq n+1 . e_j \notin \chi_k) = c_0 .$$

However, we will get an increase in domain conflict from c_0 to c_0^* when we move e_q to χ_{n+1}. This increase is an evidence indicating that e_q does not belong to χ_{n+1}, $m_{\Delta\chi}(e_q \notin \chi_{n+1})$, and the new domain conflict after e_q is moved into χ_{n+1} is interpreted as

$$m_{\chi + \{\chi_{n+1}\}}(\exists j \forall k . e_j \notin \chi_k) = c_0^* .$$

We will derive $m_{\Delta\chi}(e_q \notin \chi_{n+1})$ by stating that

$$\mathrm{Bel}_{\chi + \{\chi_{n+1}\}}(\exists j \forall k . e_j \notin \chi_k) = \mathrm{Bel}_{\Delta\chi \oplus \chi}(\exists j \forall k . e_j \notin \chi_k),$$

where

$$\mathrm{Bel}_{\chi + \{\chi_{n+1}\}}(\exists j \forall k . e_j \notin \chi_k) = c_0^*$$

and

$$\mathrm{Bel}_{\Delta\chi \oplus \chi}(\exists j \forall k . e_j \notin \chi_k) = c_0 + m_{\Delta\chi}(e_q \notin \chi_{n+1}) \cdot [1 - c_0] .$$

Thus, we received

$$m_{\Delta\chi}(e_q \notin \chi_{n+1}) = \frac{c_0^* - c_0}{1 - c_0}$$

as the sought for evidence, indicating that e_q does not belong to χ_{n+1}.

We will also receive an evidence from domain conflict variations if e_q is in a subset χ_i by itself and moved from χ_i to another already existing subset. In this case we may get either an increase or decrease in domain conflict. First, if the domain conflict decreases $c_0^* < c_0$ when we move e_q out from χ_i this is interpreted as an evidence that e_q does not belongs to χ_i,

$$m_{\Delta\chi}(e_q \notin \chi_i) = \frac{c_0 - c_0^*}{1 - c_0^*} .$$

Secondly, if we observe an increase in domain conflict $c_0^* > c_0$ we will interpret this as a new type of evidence, supporting the case that e_q does belong to χ_i;

$$m_{\Delta\chi}(e_q \in \chi_i) = \frac{c_0}{c_0^*}.$$

2.2.2. *Specifying evidences*

We may now combine the evidences from conflict variations and calculate the belief and plausibility for each subset that e_q belongs to the subset. The belief for this will always be zero, except when $e_q \in \chi_i$, $|\chi_i| = 1$ and $c_0 < c_0^*$, since every proposition with this one exception states that e_q does not belong to some subset.

When all evidences regarding e_q are combined we have received support for a proposition stating that e_q does not belong to any of the subsets and can not be put into a subset by itself. That proposition is false and its support is the conflict in Dempster's rule and an implication that the evidence is false.

For the case when e_q is in χ_i and $|\chi_i| > 1$ we combined all evidences regarding e_q and receive a new basic probability assignment with

$$\forall\chi^* . m^*(e_q \notin (\vee\chi^*)) = \frac{1}{1-k} \cdot \prod_{\chi_j \in \chi^*} m(e_q \notin \chi_j) \cdot \prod_{\chi_j \in (\chi - \chi^*)} [1 - m(e_q \notin \chi_j)]$$

where $\chi^* \in 2^\chi$, $\chi = \{\chi_1, \ldots, \chi_{n+1}\}$ and

$$k = \prod_{j=1}^{n+1} m(e_q \notin \chi_j).$$

This gave us plausibilities that e_q belongs to a subset of

$$\forall k \neq n+1 . \mathrm{Pls}(e_q \in \chi_k) = \frac{1 - m(e_q \notin \chi_k)}{1 - \prod_{j=1}^{n+1} m(e_q \notin \chi_j)}$$

and

$$\mathrm{Pls}(e_q \in \chi_{n+1}) = \frac{1 - m(e_q \notin \chi_{n+1})}{1 - \prod_{j=1}^{n+1} m(e_q \notin \chi_j)}.$$

For the situation when $e_q \in \chi_i$, $|\chi_i| = 1$ and $c_0 > c_0^*$, the only change was that the domain conflict variation appeared in the i^{th} evidence instead of the $n{+}1^{\mathrm{th}}$. This gave us a slight change in the calculation of plausibility. For subsets except χ_i we had

$$\forall k \neq i.\, \text{Pls}(e_q \in \chi_k) = \frac{1 - m(e_q \notin \chi_k)}{1 - \prod_{j=1}^{n} m(e_q \notin \chi_j)}$$

and for χ_i

$$\text{Pls}(e_q \in \chi_i) = \frac{1 - m(e_q \notin \chi_i)}{1 - \prod_{j=1}^{n} m(e_q \notin \chi_j)}.$$

When $e_q \in \chi_i$, $|\chi_i| = 1$ and $c_0 < c_0^*$ we did not receive any conflict in the combination of all evidences regarding e_q since we had no evidence against the proposition that e_q belonged to χ_i. Furthermore, when we calculate belief and plausibility for any subset other than χ_i we have a zero belief in that e_q belongs to χ_k but we receive a plausibility of

$$\forall k \neq i.\, \text{Pls}(e_q \in \chi_k) = [1 - m(e_q \in \chi_i)] \cdot [1 - m(e_q \notin \chi_k)]$$

and for χ_i we receive a belief of

$$\text{Bel}(e_q \in \chi_i) = m(e_q \in \chi_i) + [1 - m(e_q \in \chi_i)] \cdot \prod_{\chi_j \in \chi^{-i}} m(e_q \notin \chi_j)$$

where $\chi^{-i} = \chi - \{\chi_i\}$, $\chi = \{\chi_1, \ldots, \chi_n\}$ and a plausibility of one.

2.2.3. *Handling the falsity of evidences*

In combining the evidences regarding e_q in Sec. 2.2.2 we received some support k for the proposition that e_q did not belong to any of the subsets. This is impossible and implies to a degree k that e_q is a false evidence. If we had no indication as to the possible falsity of e_q we would take no action, but if there existed such an indication we would pay ever less regard to the evidence the higher the degree was that the evidence is false and pay no attention to the evidence when it is certainly false. This was done by discounting the evidence with its credibility α,

$$m^{\%}(A_j) = \begin{cases} \alpha \cdot m(A_j), & A_j \neq \Theta \\ 1 - \alpha + \alpha \cdot m(\Theta), & A_j = \Theta \end{cases},$$

where A_j is e_q or Θ, $m(\cdot)$ the original evidence, and where the credibility α is defined as one minus the support in the false proposition that e_q does not belong to any subset and cannot be put in a subset by itself, i.e. one minus the conflict in Dempster's rule when combining all evidences regarding e_q;

$$\alpha \stackrel{\Delta}{=} 1 - m^{*}(e_q \notin (\ \vee \chi)) = 1 - k.$$

2.2.4. *Finding usable evidences*

If we plan to use e_q in the reasoning process of some event, we must find the credibility that e_q belongs to the subset in question and then discount the evidence by its credibility.

Here we should note that each original piece of evidence can be used in the reasoning process of any subset that it belongs to with a plausibility above zero, given only that it is discounted to its credibility in belonging to the subset.

An evidence that cannot possible belong to a subset χ_i has a credibility of zero and should be discounted entirely for that subset, while an evidence which cannot possibly belong to any other subset χ_j and is without any support whatsoever against χ_i has a credibility of one and should not be discounted at all when used in the reasoning process for χ_i. That is, the degree to which an evidence can belong to a subset and no other subset corresponds to the importance the evidence should be allowed to play in that subset.

We derived the credibility α_j of e_q when e_q is used in χ_j as

$$\alpha_j = [1 - \mathrm{Bel}(e_q \in \chi_i)] \cdot \frac{[\mathrm{Pls}(e_q \in \chi_j)]^2}{\sum_k \mathrm{Pls}(e_q \in \chi_k)}, \qquad j \neq i,$$

$$\alpha_i = \mathrm{Bel}(e_q \in \chi_i) + [1 - \mathrm{Bel}(e_q \in \chi_i)] \cdot \frac{[\mathrm{Pls}(e_q \in \chi_i)]^2}{\sum_k \mathrm{Pls}(e_q \in \chi_k)},$$

where $\mathrm{Bel}(e_q \in \chi_i)$ is equal to zero except when $e_q \in \chi_i$, $|\chi_i| = 1$ and $c_0 < c_0^{*}$. This gave us a final discounted bpa as

$$m^{\%\%_i}(A_j) = \begin{cases} \alpha_i \cdot m^{\%}(A_j), & A_j \neq \Theta \\ 1 - \alpha_i + \alpha_i \cdot m^{\%}(\Theta), & A_j = \Theta \end{cases}.$$

3. Deriving a posterior domain probability distribution

3.1. *Evidences about subsets*

We use the idea that each evidence in a subset supports the existence of that subset to the degree that that evidence supports anything at all. For a subset χ_i, each single evidence we have is discounted for its degree of falsity and its degree of credibility in belonging to χ_i, $m_q^{\%\%i}$. All discounted evidences in χ_i are then combined. The value of all $m_q^{\%\%i}$'s were derived in [2] from the m_q's by the specifying process. The degree to which the bpa resulting from this combination supports anything at all other than the entire frame is the degree to which these evidences taken together supports the existence of χ_i, i.e. that χ_i is a nonempty subset that belongs to χ. Thus, we have

$$m_{\chi_i}(\chi_i \in \chi) = 1 - \frac{1}{1-k} \cdot \prod_q m_q^{\%\%i}(\Theta),$$

$$m_{\chi_i}(\Theta) = \frac{1}{1-k} \cdot \prod_q m_q^{\%\%i}(\Theta)$$

where k is the conflict in Dempster's rule when combining all $m_q^{\%\%i}$.

For every evidence we have some support in favor of the evidence not belonging to the subset. To the degree that this is fulfilled for all evidences in χ_i it supports the case that none of the evidences that could belong to χ_i actually did so. That is, it is support for the case that the subset is false. Thus, we would like to discount the just derived evidences as

$$m_{\chi_i}^{\%}(\chi_i \in \chi) = \alpha_i \cdot m_{\chi_i}(\chi_i \in \chi),$$

$$m_{\chi_i}^{\%}(\Theta) = 1 - \alpha_i + \alpha_i \cdot m_{\chi_i}(\Theta)$$

where

$$\alpha_i = \begin{cases} 1, & |\chi_i| = 1, c_0 < c_0^* \\ 1 - f_i \cdot g_i \cdot h_i, & \text{otherwise} \end{cases}$$

with

$$f_i = \prod_{q|\, e_q \in \chi_i} m(e_q \notin \chi_i),\ [\,(|\chi_i| > 1) \vee (|\chi_i| = 1, c_0 > c_0^*)\,],$$

$$g_i = \begin{cases} \prod_{q|\, e_q \in \chi_j, j \neq i} m(e_q \notin \chi_i), & (|\chi_i| > 0,\ [\,(|\chi_j| > 1) \vee (|\chi_j| = 1, c_0 > c_0^*)\,]) \\ \prod_{q|\, e_q \in \chi_j, j \neq n+1} m(e_q \notin \chi_{n+1}), & |\chi_{n+1}| = 0, |\chi_j| > 1 \end{cases}$$

and

$$h_i = \prod_{q|\ e_q \in \chi_j,\ j \neq i} [1 - (1 - m(e_q \notin \chi_i)) \cdot (1 - m(e_q \in \chi_j))],\ |\chi_i| > 0,\ |\chi_j| = 1,\ c_0 < c_0^*$$

where $f_i \cdot g_i \cdot h_i$ is the support that χ_i is empty, i.e. support that χ_i does not exist. Here we have, from [2],

$$\forall i,\ e_q \in \chi_i .\ m(e_q \notin \chi_j) = \begin{cases} \dfrac{c_0^* - c_0}{1 - c_0},\ j = n + 1,\ |\chi_i| > 1 \\ \dfrac{c_i - c_i^*}{1 - c_i^*},\ j = i,\ |\chi_i| > 1 \\ \dfrac{c_0 - c_0^*}{1 - c_0^*},\ j = i,\ |\chi_i| = 1,\ c_0 > c_0^* \\ \dfrac{c_j^* - c_j}{1 - c_j},\ \text{otherwise} \end{cases}$$

and

$$\forall i,\ e_q \in \chi_i .\ m(e_q \in \chi_i) = \frac{c_0}{c_0^*},\ |\chi_i| = 1,\ c_0 < c_0^*$$

where c_i and c_i^* are conflicts in subset χ_i before and after e_q is taken out from the subset, c_j^* and c_j are conflict in a subset χ_j, $j \neq i$, before and after e_q was brought into the subset, and c_0 and c_0^* are domain conflicts before and after e_q was brought either from a subset with several evidences into a new subset or, if it is in a subset by itself, from this subset into one of the other already existing subsets.

3.2. *Evidence about the number of subsets*

The discounted evidences $m_{\chi_i}^{\%}$, one from each subset, are then combined. The resulting bpa will then have focal elements that supports propositions such as

$$(\chi_1 \in \chi) \wedge (\chi_3 \in \chi) \wedge (\chi_4 \in \chi).$$

We have

$$m_{\chi}^{\%}((\ \wedge\chi^{*}\ ) \in \chi) = \prod_{i|\ (\chi_i \in \chi^{*}\ )} m_{\chi_i}^{\%}(\chi_i \in \chi) \cdot \prod_{j|\ (\chi_j \notin \chi^{*}\ )} m_{\chi_j}^{\%}(\Theta),$$

$$m_{\chi}^{\%}(\Theta) = \prod_{i=1}^{n} m_{\chi_i}^{\%}(\Theta).$$

From this we can create a new bpa by exchanging all propositions in the previous bpa that are conjunctions of r terms for one proposition in the new bpa that is on the form $|\chi| \geq r$. The sum of probability of all conjunctions of length r in the previous bpa is then awarded the focal element in the new bpa which supports the proposition that $|\chi| \geq r$;

$$m_{\chi}(|\chi| \geq r) = \sum_{\chi^{*}|\ |\chi^{*}| = r} m_{\chi}^{\%}((\ \wedge\chi^{*}\ ) \in \chi),$$

$$m_{\chi}(\Theta) = m_{\chi}^{\%}(\Theta)$$

where $\chi^{*} \in 2^{\chi}$ and $\chi = \{\chi_1, \chi_2, \ldots, \chi_n\}$.

A proposition in the new bpa is then a statement about the existence of some minimal number of subsets and its bpa taken as a whole gives us an opinion about the probability of different numbers of subsets.

3.3. *Combining the evidence with a prior distribution*

This newly created bpa can now be combined with our prior probability distribution, $m(\cdot)$, from the problem specification, to yield the demanded posterior probability distribution, $m^{*}(\cdot)$. We get

$$m^{*}(E_i) = \frac{1}{1-k} \cdot m(E_i) \cdot \left(m_{\chi}(\Theta) + \sum_{j=1}^{i} m_{\chi}(|\chi| \geq j)\right)$$

where

$$k = \sum_{i=0}^{n-1} \sum_{j=i+1}^{n} m(E_i) \cdot m_{\chi}(|\chi| \geq j)$$

is the conflict in that final combination.

Thus, by viewing each evidence in a subset as support for the existence of that subset we were able to derive a bpa, concerned with the question of how many subsets we have, which we could combine with our prior domain probability distribution in order to obtain the sought-after posterior domain probability distribution.

4. An Example

In our first article [1] we described a problem involving two possible burglaries. In this example we had evidence weakly specified in the sense that it is uncertain to which possible burglary the propositions are referring. The problem we were facing was described as follows:

> *Assume that a baker's shop at One Baker Street has been burglarized, event* 1. *Let there also be some indication that a baker's shop across the street, at Two Baker Street, might have been burglarized, although no burglary has been reported, event* 2. *An experienced investigator estimates that a burglary has taken place at Two Baker Street with a probability of* 0.4. *We have received the following evidences. A credible witness reports that "a brown-haired man who is not an employee at the baker's shop committed the burglary at One Baker Street," evidence* 1. *An anonymous witness, not being aware that there might be two burglaries, has reported "a brown-haired man who works at the baker's shop committed the burglary at Baker Street," evidence* 2. *Thirdly, a witness reports having seen "a suspicious-looking red-haired man in the baker's shop at Two Baker Street," evidence* 3. *Finally, we have a fourth witness, this witness, also anonymous and not being aware of the possibility of two burglaries, reporting that the burglar at the Baker Street baker's shop was a brown-haired man. That is, for example:*

evidence 1*:*
  proposition:
    action part: BO
    event part: E_1*:*
$m(BO) = 0.8$
$m(\Theta) = 0.2$

evidence 2*:*
  proposition:
    action part: BI
    event part: E_1, E_2
$m(BI) = 0.7$
$m(\Theta) = 0.3$

evidence 3*:*
  proposition:
    action part: R
    event part: E_2*:*
$m(R) = 0.6$
$m(\Theta) = 0.4$

evidence 4*:*
  proposition:
    action part: B
    event part: E_1, E_2
$m(B) = 0.5$
$m(\Theta) = 0.5$

domain probability distribution:

$$m(E_i) = \begin{cases} 0.6, & i = 1 \\ 0.4, & i = 2 \\ 0, & i \neq 1, 2 \end{cases} .$$

All evidences were originally put into one subset, χ_1. By minimizing the metaconflict function it was found best to partition the evidences into two subsets. The minimum of the metaconflict function was found when evidences one and four were moved from χ_1 into χ_2 while evidences two and three remained in χ_1. This gave us a conflict in χ_1 of $c_1 = 0.42$, in χ_2 of $c_2 = 0$, and a domain conflict of $c_0 = 0.6$.

In our second article [2] we studied variations in the cluster conflict when a piece of evidence is moved from one subset to another, or put into a new subset by itself. Starting with e_1 we found that if e_1 in χ_2 is moved out from χ_2 the conflict remains at zero, $c_2^* = 0$. If e_1 then is moved into χ_1 its conflict increased to $c_1^* = 0.788$, but if e_1 is instead put into a subset by itself, χ_3, we will have a domain conflict of one, $c_0^* = 1$. By the formulas of [2] we received three bpa's regarding e_1:

$$m(e_1 \notin \chi_1) = \frac{c_1^* - c_1}{1 - c_1} = 0.634, \qquad m(e_1 \notin \chi_2) = \frac{c_2 - c_2^*}{1 - c_2^*} = 0$$

and

$$m(e_1 \notin \chi_3) = \frac{c_0^* - c_0}{1 - c_0} = 1,$$

with the remainder in each case awarded to the entire frame. We received for the other three evidences by the same formulas:

$$m(e_2 \notin \chi_i) = \begin{cases} 0.42, i = 1 \\ 0.56, i = 2 \\ 1, i = 3 \end{cases}, \qquad m(e_3 \notin \chi_i) = \begin{cases} 0.42, i = 1 \\ 0.54, i = 2 \\ 1, i = 3 \end{cases},$$

$$m(e_4 \notin \chi_i) = \begin{cases} 0.155, i = 1 \\ 0, i = 2 \\ 1, i = 3 \end{cases}.$$

In each case the remainder was awarded to the entire frame.

When the three bpa's regarding where a particular evidence might belong were combined, a conflict was received for e_2 and e_3, but not for e_1 and e_4. Thus, there is no indication from this combination that e_1 and e_4 might be false. For the second and third evidence a conflict of 0.2352 and 0.2268 was received, respectively. This is their degrees of falsity. Evidences e_2 and e_3 were then discounted to their respective degrees of credibility $\alpha = 1 - k$, i.e. 0.7648 and 0.7732:

$$m^{\%}(A_j) = \begin{cases} \alpha \cdot m(A_j), & A_j \neq \Theta \\ 1 - \alpha + \alpha \cdot m(\Theta), & A_j = \Theta \end{cases}.$$

This gave us

$$m_1^{\%}(BO) = 0.8 \qquad m_2^{\%}(BI) = 0.5354 \qquad m_3^{\%}(R) = 0.4639 \qquad m_4^{\%}(B) = 0.5$$

$$m_1^{\%}(\Theta) = 0.2 \qquad m_2^{\%}(\Theta) = 0.4646 \qquad m_3^{\%}(\Theta) = 0.5361 \qquad m_4^{\%}(\Theta) = 0.5.$$

Since all four evidences can belong to either of the two subsets it will always be uncertain if it belongs to a particular subset in question. In order to justify the use of an evidence in some subset we must find the credibility that it belongs to the subset and discount the evidence to its credibility. That is, an individual discounting is made for each subset and evidence according to how credible it is that the evidence belongs to the subset.

The credibility that e_1 belongs to χ_1 is

$$\alpha_1 = \frac{(\mathrm{Pls}(e_1 \in \chi_1))^2}{\sum_{j=1}^{2} \mathrm{Pls}(e_1 \in \chi_j)} = \frac{0.366^2}{0.366 + 1} = 0.0981$$

where

$$\mathrm{Pls}(e_1 \in \chi_1) = \frac{1 - m(e_1 \notin \chi_1)}{1 - m(e_1 \notin \chi_1) \cdot m(e_1 \notin \chi_2) \cdot m(e_1 \notin \chi_3)} = 1 - 0.634 = 0.366,$$

$$\mathrm{Pls}(e_1 \in \chi_2) = \frac{1 - m(e_1 \notin \chi_2)}{1 - m(e_1 \notin \chi_1) \cdot m(e_1 \notin \chi_2) \cdot m(e_1 \notin \chi_3)} = 1.$$

and that e_1 belongs to χ_2

$$\alpha_2 = \frac{(\mathrm{Pls}(e_1 \in \chi_2))^2}{\sum_{j=1}^{2} \mathrm{Pls}(e_1 \in \chi_j)} = \frac{1}{0.366 + 1} = 0.7321.$$

For the other three evidences we get: e_2: $\alpha_1 = 0.4310$, $\alpha_2 = 0.2480$, e_3: $\alpha_1 = 0.4182$, $\alpha_2 = 0.2632$, and for e_4: $\alpha_1 = 0.3870$, $\alpha_2 = 0.5420$.

Discounting the four evidences to their credibility of belonging to χ_1 and χ_2, respectively, we found:

$$m_1^{\%\%1}(BO) = 0.0784 \qquad m_1^{\%\%2}(BO) = 0.5856$$

$$m_1^{\%\%1}(\Theta) = 0.9216 \qquad m_1^{\%\%2}(\Theta) = 0.4144,$$

$$m_2^{\%\%1}(BI) = 0.2308 \qquad m_2^{\%\%2}(BI) = 0.1328$$

$$m_2^{\%\%1}(\Theta) = 0.7692 \qquad m_2^{\%\%2}(\Theta) = 0.8672,$$

$$m_3^{\%\%1}(R) = 0.1940 \qquad m_3^{\%\%2}(R) = 0.1221$$

$$m_3^{\%\%1}(\Theta) = 0.8060 \qquad m_3^{\%\%2}(\Theta) = 0.8779$$

and

$$m_4^{\%\%1}(B) = 0.1935 \qquad m_4^{\%\%2}(B) = 0.2710$$

$$m_4^{\%\%1}(\Theta) = 0.8065 \qquad m_4^{\%\%2}(\Theta) = 0.7290.$$

These results were derived in [2].

Starting with these results we begin the work to find a posterior probability distribution for the number of subsets.

By using the idea that each evidence in a subset supports the existence of that subset to the degree that the evidence supports anything at all, we calculate the support in our two subsets as

$$m_{\chi_1}(\chi_1 \in \chi) = 1 - \prod_q m_q^{\%\%1}(\Theta) = 1 - m_1^{\%\%1}(\Theta) \cdot m_2^{\%\%1}(\Theta) \cdot m_3^{\%\%1}(\Theta) \cdot m_4^{\%\%1}(\Theta) = 0.4893,$$

$$m_{\chi_1}(\Theta) = \prod_q m_q^{\%\%1}(\Theta) = m_1^{\%\%1}(\Theta) \cdot m_2^{\%\%1}(\Theta) \cdot m_3^{\%\%1}(\Theta) \cdot m_4^{\%\%1}(\Theta) = 0.5107$$

and

$$m_{\chi_2}(\chi_2 \in \chi) = 1 - \prod_q m_q^{\%\%2}(\Theta) = 1 - m_1^{\%\%2}(\Theta) \cdot m_2^{\%\%2}(\Theta) \cdot m_3^{\%\%2}(\Theta) \cdot m_4^{\%\%2}(\Theta) = 0.7268,$$

$$m_{\chi_2}(\Theta) = \prod_q m_q^{\%\%2}(\Theta) = m_1^{\%\%2}(\Theta) \cdot m_2^{\%\%2}(\Theta) \cdot m_3^{\%\%2}(\Theta) \cdot m_4^{\%\%2}(\Theta) = 0.2732.$$

If we have support for every single piece of evidence in some subset in favor of that the evidence does not belong to the subset, then this is also support that the subset is false. In this case none of the evidences that could belong to the subset actually did so and the subset was derived by mistake. Thus, we will discount the just derived evidences that support the existence of the subsets for this possibility.

There is some evidence against the first subset, yielding a credibility for that subset of less than one

$$\alpha_1 = 1 - \prod_{q|\, e_q \in \chi_1} m(e_q \in \chi_1) \cdot \prod_{q|\, e_q \in \chi_j,\, j \notin 1} m(e_q \in \chi_1) = 0.9826,$$

$$\alpha_2 = 1 - \prod_{q|\, e_q \in \chi_2} m(e_q \in \chi_2) \cdot \prod_{q|\, e_q \in \chi_j,\, j \notin 2} m(e_q \in \chi_2) = 1.$$

We then discount the two bpa's that support the existence of the subsets to their respective credibility and receive

$$m^{\%}_{\chi_1}(\chi_1 \in \chi) = \alpha \cdot m_{\chi_1}(\chi_1 \in \chi) = 0.4808,$$

$$m^{\%}_{\chi_1}(\Theta) = 1 - \alpha - \alpha \cdot m_{\chi_1}(\Theta) = 0.5192$$

for the first subset and

$$m^{\%}_{\chi_2}(\chi_2 \in \chi) = \alpha \cdot m_{\chi_1}(\chi_2 \in \chi) = 0.7268,$$

$$m^{\%}_{\chi_2}(\Theta) = 1 - \alpha - \alpha \cdot m_{\chi_1}(\Theta) = 0.2732$$

for the second subset. If we then combine these two bpa's we receive

$$m^{\%}_{\chi}((\ \wedge \{\chi_1, \chi_2\}) \in \chi) = \prod_{i|\ (\chi_i \in \{\chi_1, \chi_2\})} m^{\%}_{\chi_i}(\chi_i \in \chi) \cdot \prod_{j|\ (\chi_j \notin \{\chi_1, \chi_2\})} m^{\%}_{\chi_j}(\Theta)$$

$$= m^{\%}_{\chi_1}(\chi_1 \in \chi) \cdot m^{\%}_{\chi_2}(\chi_2 \in \chi) = 0.3494,$$

$$m^{\%}_{\chi}((\ \wedge \{\chi_1\}) \in \chi) = \prod_{i|\ (\chi_i \in \{\chi_1\})} m^{\%}_{\chi_i}(\chi_i \in \chi) \cdot \prod_{j|\ (\chi_j \notin \{\chi_1\})} m^{\%}_{\chi_j}(\Theta)$$

$$= m^{\%}_{\chi_1}(\chi_1 \in \chi) \cdot m^{\%}_{\chi_2}(\Theta) = 0.1314,$$

$$m^{\%}_{\chi}((\ \wedge \{\chi_2\}) \in \chi) = \prod_{i|\ (\chi_i \in \{\chi_2\})} m^{\%}_{\chi_i}(\chi_i \in \chi) \cdot \prod_{j|\ (\chi_j \notin \{\chi_2\})} m^{\%}_{\chi_j}(\Theta)$$

$$= m^{\%}_{\chi_2}(\chi_2 \in \chi) \cdot m^{\%}_{\chi_1}(\Theta) = 0.3774,$$

$$m^{\%}_{\chi}(\Theta) = \prod_{i=1}^{2} m^{\%}_{\chi_i}(\Theta) = m^{\%}_{\chi_1}(\Theta) \cdot m^{\%}_{\chi_2}(\Theta) = 0.1418.$$

Given this result we create a new and final bpa by exchanging the focal elements of this bpa. Where the previous bpa is concerned with the question of which subsets have support, the new bpa is concerned with the question of how many subsets are supported. Thus, the new bpa gives us an opinion, based only on the result of the evidence specifying process, about the probability of different number of subsets. We have

$$m_\chi(|\chi| \geq 2) = \sum_{\chi^* \mid |\chi^*| = 2} m_\chi^{\%}((\ \wedge \chi^*) \in \chi) = m_\chi^{\%}((\ \wedge \{\chi_1, \chi_2\}) \in \chi) = 0.3494,$$

$$m_\chi(|\chi| \geq 1) = \sum_{\chi^* \mid |\chi^*| = 1} m_\chi^{\%}((\ \wedge \chi^*) \in \chi) = m_\chi^{\%}((\ \wedge \{\chi_1\}) \in \chi) + m_\chi^{\%}((\ \wedge \{\chi_2\}) \in \chi)$$

$$= 0.5087,$$

$$m_\chi(\Theta) = m_\chi^{\%}(\Theta) = 0.1418.$$

To conclude the analysis we combine this final bpa, from the evidence specifying process, with the given prior domain probability distribution from the problem specification,

$$m(E_i) = \begin{cases} 0.6, i = 1 \\ 0.4, i = 2 \\ 0, \ \text{otherwise} \end{cases},$$

in order to receive the sought-after posterior domain distribution as the bpa of that combination. When doing this we receive a conflict of

$$k = \sum_{i=0}^{1} \sum_{j=i+1}^{2} m(E_i) \cdot m_\chi(|\chi| \geq j) = m(E_0) \cdot m_\chi(|\chi| \geq 1) + m(E_0) \cdot m_\chi(|\chi| \geq 1)$$

$$+ m(E_1) \cdot m_\chi(|\chi| \geq 2) = 0.2097$$

and obtain

$$m^*(E_2) = \frac{1}{1-k} \cdot m(E_2) \cdot \left(m_\chi(\Theta) + \sum_{j=1}^{2} m_\chi(|\chi| \geq j) \right)$$

$$= \frac{1}{1-k} \cdot m(E_2) \cdot (m_\chi(\Theta) + m_\chi(|\chi| \geq 1) + m_\chi(|\chi| \geq 2)) = 0.5061,$$

$$m^*(E_1) = \frac{1}{1-k} \cdot m(E_1) \cdot \left(m_\chi(\Theta) + \sum_{j=1}^{1} m_\chi(|\chi| \geq j) \right)$$

$$= \frac{1}{1-k} \cdot m(E_2) \cdot (m_\chi(\Theta) + m_\chi(|\chi| \geq 1)) = 0.4939,$$

$$m^*(E_i) = 0, \ \text{otherwise}$$

as the posterior domain probability distribution. We find from the posterior distribution that the alternative with two events is slightly preferable to the one-event alternative.

5. Conclusions

We have shown that it is possible to derive a posterior domain probability distribution from the reasoning process of specifying nonspecific evidence. This was done by viewing each evidence in a subset as support for the existence of that subset. Based on this, we were able to find support for different number of subsets. Combined with a given prior distribution that yielded the sought-after posterior distribution.

The methodology described in this article builds on the work to partition the set of evidences by minimizing a criterion function of overall conflict that was established within Dempster-Shafer theory [1] and also on the work of specifying evidences by studying changes in the conflict when a piece of evidence was moved from one subset to another [2].

Acknowledgments

I would like to thank Stefan Arnborg, Ulla Bergsten, and Per Svensson for their helpful comments regarding this article.

References

1. J. Schubert, "On nonspecific evidence", *Int. J. Intell. Syst.* **8** (1993) 711-725.
2. J. Schubert, "Specifying nonspecific evidence", FOA Report C 20975-2.7, National Defence Research Establishment, Stockholm, 1994.
3. A.P. Dempster, "A generalization of Bayesian inference", *J. R. Stat. Soc. Ser. B* **30** (1968) 205-247.
4. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University, Princeton, 1976).
5. G. Shafer, "Perspectives on the Theory and Practice of Belief Functions", *Int. J. Approx. Reasoning* **4** (1990) 323-362.
6. G. Shafer, "Rejoinders to comments on 'Perspectives on the Theory and Practice of Belief Functions'", *Int. J. Approx. Reasoning* **6** (1992) 445-480.
7. U. Bergsten and J. Schubert, "Dempster's rule for evidence ordered in a complete directed acyclic graph", *Int. J. Approx. Reasoning* **9** (1993) 37-73.
8. J.D. Lowrance and T.D. Garvey, "Evidential Reasoning: An Implementation for Multisensor Integration", Technical Note 307, SRI International, Menlo Park, CA, 1983.

Dempster's Rule for evidence Ordered in a Complete directed Acyclic Graph

Int. J. Approx . Reasoning, **9**(1), 37-73, 1993.

IV

Dempster's Rule for Evidence Ordered in a Complete Directed Acyclic Graph

Ulla Bergsten and Johan Schubert
Division of Applied Mathematics and Scientific Data Processing, Department of Weapon Systems, Effects and Protection, National Defence Research Establishment, Sundbyberg, Sweden.

ABSTRACT

For the case of evidence ordered in a complete directed acyclic graph this paper presents a new algorithm with lower computational complexity for Dempster's rule than that of step-by-step application of Dempster's rule. In this problem, every original pair of evidences, has a corresponding evidence against the simultaneous belief in both propositions. In this case, it is uncertain whether the propositions of any two evidences are in logical conflict. The original evidences are associated with the vertices and the additional evidences are associated with the edges. The original evidences are ordered, i.e., for every pair of evidences it is determinable which of the two evidences is the earlier one. We are interested in finding the most probable completely specified path through the graph, where transitions are possible only from lower- to higher-ranked vertices. The path is here a representation for a sequence of states, for instance a sequence of snapshots of a physical object's track. A completely specified path means that the path includes no other vertices than those stated in the path representation, as opposed to an incompletely specified path that may also include other vertices than those stated. In a hierarchical network of all subsets of the frame, i.e., of all incompletely specified paths, the original and additional evidences support subsets that are not disjoint, thus it is not possible to prune the network to a tree. Instead of propagating belief, the new algorithm reasons about the logical conditions of a completely specified path through the graph. The new algorithm is $O(|\Theta| \log|\Theta|)$*, compared to* $O(|\Theta|^{\log|\Theta|})$ *of the classic brute force algorithm. After a detailed presentation of the reasoning behind*

Address correspondence to Johan Schubert, Division of Applied Mathematics and Scientific Data Processing, Department of Weapon Systems, Effects and Protection, National Defence Research Establishment, S-172 90 Sundbyberg, Sweden.
Received May, 1991; accepted December 5, 1992

International Journal of Approximate Reasoning 1993; 9:37–73

655 Avenue of the Americas, New York, NY 10010 0888-613X/93/$6.00

the new algorithm we conclude that it is feasible to reason without approximation about completely specified paths through a complete directed acyclic graph.

KEYWORDS: ***belief functions, Dempster–Shafer theory, Dempster's rule, evidential reasoning, uncertainty, propagation of evidence, hierarchical network, partitions, complete directed acyclic graph, computational complexity***

1. INTRODUCTION

The development of knowledge-based systems has evoked increasing attention to the subject of approximate reasoning. The available information in a system is often uncertain, incomplete, and even partly incorrect—demanding methods able to handle this kind of information. The Dempster–Shafer theory, which provides an attractive representation of uncertainty and an intuitive combination of uncertain information, is one such method (Dempster [1], Shafer [2, 3, 4]). However, one problem with the Dempster–Shafer theory is its computational complexity. In many cases even a moderate amount of data leads to huge computational complexity making it necessary either to aggregate focal elements, i.e., use summarization (Lowrance et al. [5]), or to derive approximate or special case algorithms.

In this paper we present an algorithm for the special case of evidences ordered in a complete directed acyclic graph. In this case, it is uncertain whether the propositions of any two evidences are in logical conflict. Here, we can model the uncertainty by an additional evidence against the simultaneous belief in both propositions and treat the two original propositions as non-conflicting. This will give rise to a complete directed acyclic graph with the original evidences on the vertices and the additional ones on the edges. As an example, we may think of the vertices as positions in time and space and the edges as transitions between these positions. Transitions are only possible from a vertex with a lower index to one with a higher. We are interested in finding the most probable path of an object. The evidence at a vertex may then be an evidence that the object has been at that position and the evidence at an edge an evidence against the possibility of a transition between the two positions. The classic algorithm calculates the support and plausibility for a given path, i.e., a sequence of vertices, through the graph by first combining all evidences step-by-step with Dempster's rule and then summing up all contributions for the path. The new algorithm reasons instead about the logical conditions of a completely specified path through the complete directed acyclic graph, gaining significantly in time and space complexity.

In this paper, we give a brief summary of Dempster–Shafer theory (Section 2), discuss the type of problem domains that satisfy our restriction, and then describe the representation of Dempster–Shafer theory in this case (Section 3). In Section 4 we review some previous work on belief propagation and compare these results to ours. We discuss how the classic algorithm works in this case and give an example (Section 5). We then give an explanation of the reasoning behind the new algorithm as well as a presentation of the formal structure of the new algorithm (Section 6). Finally, we discuss its computational complexity (Section 7).

2. DEMPSTER–SHAFER THEORY

In Dempster–Shafer theory, belief is assigned to a proposition by a basic probability assignment. The proposition is represented by a subset A of an exhaustive set of mutually exclusive possibilities, a frame of discernment Θ.

The basic probability assignment is a function from the power set of Θ to [0, 1]

$$m : 2^{\Theta} \to [0, 1]$$

whenever

$$m(\emptyset) = 0$$

and

$$\sum_{A \subseteq \Theta} m(A) = 1$$

where m(A) is called a basic probability number, that is the belief committed exactly to A.

The total belief of a proposition A is obtained from the sum of probabilities for those propositions that are subsets of the proposition in question and the probability committed exactly to A

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B)$$

where Bel(A) is the total belief in A and Bel($\cdot$) is called a belief function

$$\mathrm{Bel} : 2^{\Theta} \to [0, 1].$$

A subset A of Θ is called a focal element of Bel if the basic probability number for A is non-zero.

In addition to the belief in a proposition A it is also of interest to know how plausible a proposition might be, i.e., the degree to which we do not doubt A. The plausibility,

$$\mathrm{Pls} : 2^{\Theta} \to [0, 1]$$

is defined as

$$\mathrm{Pls}(A) = 1 - \mathrm{Bel}(A^c).$$

We can calculate the plausibility directly from the basic probability assignment

$$\mathrm{Pls}(A) = \sum_{B \cap A \neq \varnothing} m(B).$$

Thus, while belief in A measures the total probability certainly committed to A, plausibility measures the total probability that is in or can be moved into A, i.e., $\mathrm{Bel}(A) \leq \mathrm{Pls}(A)$.

If we receive a second item of information concerning the same issue from a different source, the two items can be combined to yield a more informed view. Combining two belief functions is done by calculating the orthogonal combination with Dempster's rule. This is most simply illustrated through the combination of basic probability assignments. Let A_i be a focal element of Bel_1 and let B_j be a focal element of Bel_2. Combining the corresponding basic probability assignments m_1 and m_2 results in a new basic probability assignment $\mathrm{m}_1 \oplus \mathrm{m}_2$

$$m_1 \oplus m_2(A) = \mathrm{K} \cdot \sum_{A_i \cap B_j = A} m_1(A_i) \cdot m_2(B_j)$$

where K is a normalizing constant

$$\mathrm{K} = \left(1 - \sum_{A_i \cap B_j = \varnothing} m_1(A_i) \cdot m_2(B_j)\right)^{-1}.$$

This normalization is needed because, by definition, no probability mass may be committed to $\varnothing$. The new belief function $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2(\cdot)$ can be calculated by the above formula from $\mathrm{m}_1 \oplus \mathrm{m}_2(\cdot)$.

When we wish to combine several belief functions this is simply done by combining the first two and then combine the result with the third and so forth.

3. DISCUSSION OF PROBLEM DOMAINS

3.1. Problem Domains that Satisfy the Assumptions of the Algorithm

The algorithm presented in this paper is a special case algorithm for evidences ordered in a complete directed acyclic graph, where the vertices represent states and the edges transitions between states. We are interested in finding through which sequence of states a process has developed.

At every vertex we have evidence supporting the proposition that this vertex is included in the sequence and at every edge evidence expressing the degree of doubt about a transition between the corresponding states.

As an example we may consider a graph where a state represents a point in time and space and the sequence of states represents a path along which some object may have moved. For some coordinates we have evidences whose proposition tells us that this geographical point has been passed by the object at a certain time. The graph consists only of coordinates for which there is evidence. The propositions of the evidences on the edges may, for example, tell us that the time difference between the states may be too small in relation to their distance. Of course, it is impossible to move from a vertex to a previous one. There may also exist other domain-specific restrictions on the edges.

Here, we are making the assumption that only one path at a time is permitted through the graph, i.e., two objects cannot pass through the graph at the same time. The problem of analyzing paths of multiple objects can be solved by partitioning the evidences into clusters (Schubert [6]), each cluster representing a separate object, after which the problem may be solved separately for each partitioning.

The new algorithm was developed for an anti-submarine intelligence analysis system (Bergsten et al. [7]). In this application information about foreign submarine activity derives from visual observations and military sensor signals. The information is of varying quality with considerable uncertainty. Visual observations may include anything from a civilian reporting unusual wave movements on the surface to a group of naval officers recognizing a submarine tower. In shallow waters sensors may have difficulty in discerning a target, and there may be several targets present simultaneously. Thus, a non-firing sensor does not necessarily exclude a passage. Weather, wind, and water temperature are other important factors determining the range and detection probability of a sensor. From this it follows that an unknown number of observations may be false, i.e., not arising from submarines.

We are interested in finding the path along which the suspected submarine has moved, i.e., which observations are true. The problem we are treating here is simplified by the assumption that all observations arise from only one submarine.

This problem may be described by the complete directed acyclic graph discussed above. Each observation at a vertex, whether visual or originating from a sensor, is an evidence indicating that a submarine has visited the point of the observation. The vertices are ordered according to the time of the observations. Evidences at the edges, against transitions between the observations, appears as a lack of sensor signals, unrealistic velocity requirements, etc.

In this case we often have less than ten interesting observations during a certain period. This is because the incoming flow of observations is rather small, and observations soon become too old to give valuable information about the current position of the submarine.

Even with this moderate number of observations, the computational complexity becomes too high for the classic algorithm to be used, but is acceptable for the new algorithm.

3.2. Evidential Reasoning in a Complete Directed Acyclic Graph

Let a complete directed acyclic graph G be given. We are interested in transitions between vertices and search for the most probable path through the graph. Every vertex v_i in G is associated with an evidence e_i which to the amount p_i supports the proposition that this vertex belongs to the sought path S. Furthermore, for every pair of vertices v_i and v_j, there is an edge between the vertices that is associated with an evidence e_{ij} which to the amount q_{ij} speaks against a direct transition between these two vertices. Thus, e_{ij} supports the proposition that there is no transition between the vertices v_i and v_j that does not involve any other vertex between them (Figure 1). All the corresponding belief functions are simple support functions. Because the directed acyclic graph is complete, the set of vertices is totally ordered. All evidences are supposed to be independent.

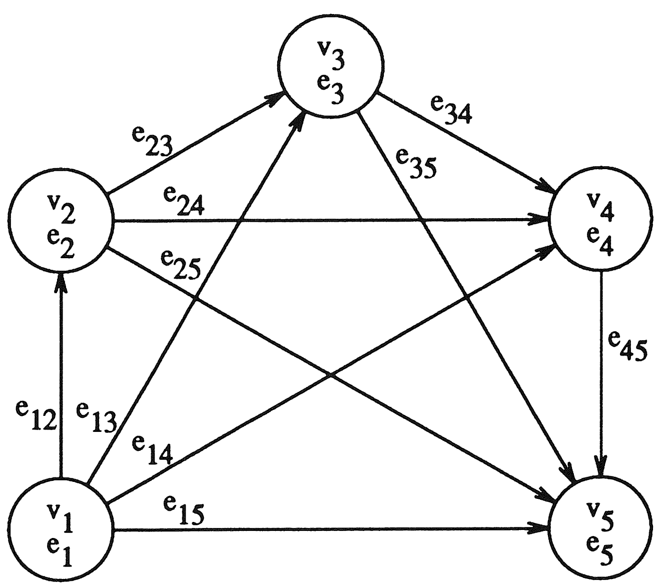

Figure 1. Evidences in the complete directed acyclic graph.

The first step in applying evidential reasoning to a given problem is to delimit the propositional space of possible situations, i.e., "the frame of discernment." In our case the frame of discernment is the set of all possible paths through the graph, where transitions are possible only from lower- to higher-ranked vertices. Assuming the graph G consists of n vertices, any path S of the frame through G can be represented by $< x_1, x_2, \ldots, x_n >$ where the i:th element corresponds to vertex v_i and takes the value r_i or $\neg r_i$ according to whether or not it is contained in this particular path. Our frame Θ will then consist of these 2^n different paths. Consider for example a graph consisting of five vertices $v_1, \ldots, v_5$ and directed edges from every vertex to all vertices with a higher index. The path from v_1 to v_4 to v_5, not including v_2 or v_3, is represented by $< r_1, \neg r_2, \neg r_3, r_4, r_5 >$. To be able to express subsets of Θ in a convenient way, we extend the range of x_i with the value θ_i, meaning either r_i or $\neg r_i$. E.g., $< r_1, \theta_2, \theta_3, r_4, r_5 >$, an incompletely specified path, will denote all paths passing through v_1, v_4, and v_5.

4. PREVIOUS WORK

There has been some work on generally applicable improvements of the time complexity of Dempster's rule, e.g. [8, 9], reducing the time complexity in the general case from $O(3^{|\Theta|})$ to $O(|\Theta| \cdot 2^{|\Theta|})$. However, most improvements have concerned important special cases. Foremost among these are methods dealing with belief propagation in trees.

4.1. Belief Propagation in Hierarchies

In 1985 Gordon and Shortliffe [10] suggested that when evidence supports singletons or disjoint subsets of the frame, a hierarchical network of subsets could be pruned to a hierarchical tree. The assumption is that a strict hierarchy of hypotheses can be defined from some subsets of 2^Θ and that a system will only receive information for these subsets. They proposed a method partly based on the work of Barnett [11] for reasoning about hypotheses with hierarchical relationships.

Barnett showed that simple support functions focused on singletons or their complements can be combined with a time complexity, for each considered subset of Θ, that is linear in the size of the frame, $|\Theta|$. In order to obtain linear time complexity, it is assumed that simple support functions with the same focus have already been combined.

Barnett's method can be described as first combining all simple support functions with equal foci and then, for each singleton, combining the resulting simple support functions for and against the singleton. For each singleton, this results in a separable support function with three focal elements: the singleton, its complement, and Θ. Finally, the separable support functions are combined separately for each considered subset of Θ in such a way that a linear time complexity is obtained. Barnett's technique will also work when the simple support functions are focused on subsets or their complements if all subsets considered are disjoint.

Besides the assumption that the domain allows a hierarchical network to be pruned to a hierarchical tree and that a system will only receive information about those subsets of the frame that are in the tree, the method by Gordon and Shortliffe is approximate in that it does not assign belief to subsets that are not in the tree. This approximation changes the time complexity from exponential to linear.

The first step is borrowed from Barnett's method. All evidences with equal foci, confirming and disconfirming, are combined, with the only difference that what Barnett did with simple support functions focused on singletons is done here for all subsets of the frame that are in the tree, T. Now there are two bpa's for each subset of the frame that is in the tree, one confirming the subset and one disconfirming it; we want to combine all bpa's in the entire tree. However, combining bpa's where some focal elements are complements of subsets in the tree might produce an intersection that is not a subset or a complement of a subset that is in the tree. We begin with the confirming bpa's. These are easily combined because the intersection between two focal elements is either empty or the smaller of the two sets. This is because of the tree structure where the focal element of a child is a subset of the focal element of the parent and where focal elements at different branches are disjoint. Finally, the disconfirming bpa's are combined one by one with m_T, where m_T is the result of the combination of all confirming bpa's. When belief is assigned to a subset, X, that is not in the tree this belief is reassigned to the smallest subset, A_i, such that X is a proper subset of A_i, $X \subset A_i$.

Shafer and Logan [12] improved on the method by Gordon and Shortliffe. They showed that, although the algorithm by Gordon and Shortliffe usually produced a good approximation its performance was not as good when used with highly conflicting evidence. Besides not being approximate, the algorithm by Shafer and Logan also calculates belief for A_i^c of every partition, A_i, that is in the tree, thus it calculates the plausibility for all partitions in the tree. Both algorithms run in linear time. Interestingly, Shafer and Logan showed that the linear time complexity of their algorithm is linear in the number of the nonterminal nodes due to the local computations of their algorithm and linear in the tree's branching factor due to Barnett's approach.

The algorithm by Shafer and Logan can handle evidence and calculate belief in partitions of the form $\{A_i, A_i^c\}$ for all subsets, A_i, in the tree. It can also calculate belief in partitions of the form $C_{A_i} \cup \{A_i^c\}$, where C_{A_i} is the set of children of A_i. However, their algorithm can not handle evidence for $C_{A_i} \cup \{A_i^c\}$. Because these two types of evidence correspond to data and domain knowledge respectively, this is a significant restriction. A generalization of the algorithm by Shafer and Logan that manages to take domain knowledge into account is the method for belief propagation in qualitative Markov trees by Shafer, Shenoy, and Mellouli [13]. In a qualitative Markov tree the children are qualitatively conditionally independent [14] given the parent, i.e., in determining which element of a child is true, there is no additional information in knowing which element of another child is true once we know which element of the parent is true. Qualitative Markov trees can arise through constructing what Shafer, Shenoy, and Mellouli call the tree of families and dichotomies. This is simply done by substituting each nonterminal node with subset A_i in a hierarchical tree by a parent-child pair with the dichotomy $\{A_i, A_i^c\}$ as subset at the parent and the family $C_{A_i} \cap \{A_i^c\}$ as subset at the child and furthermore substituting terminal nodes with subset A_i with the dichotomy $\{A_i, A_i^c\}$.

In [15] Shenoy and Shafer list the axioms under which local computations at the nodes are possible.

Shafer, Shenoy, and Mellouli point out that this computational scheme reduces the time complexity from being exponential in the size of the frame to being exponential in the size of the largest partition.

4.2. Comparison with our Method

Barnett [11] showed that it is possible to implement Dempster's rule with a time complexity linear in the size of the frame, $|\Theta|$, when the belief functions being combined are all simple support functions focused on singletons or their complements. In our case, however, the simple support functions are never focused on singletons and, with one exception, not focused on the complements of singletons. Our frame consists of all possible single paths in a complete directed acyclic graph, and the simple support functions are on subsets representing individual vertices in the complete directed acyclic graph or on subsets representing the direct transition between two vertices, i.e., on elements of 2^Θ that are not singletons or, with the exception of the two vertex graph, their complements.

Gordon and Shortliffe suggest that when evidences support singletons or disjoint subsets of the frame the hierarchical network of subsets could be pruned to a tree. Then they suggested methods for the combination of evidence in trees. Our case can of course also be represented with a hierarchical network of subsets, as seen by the example in Figure 2 of a

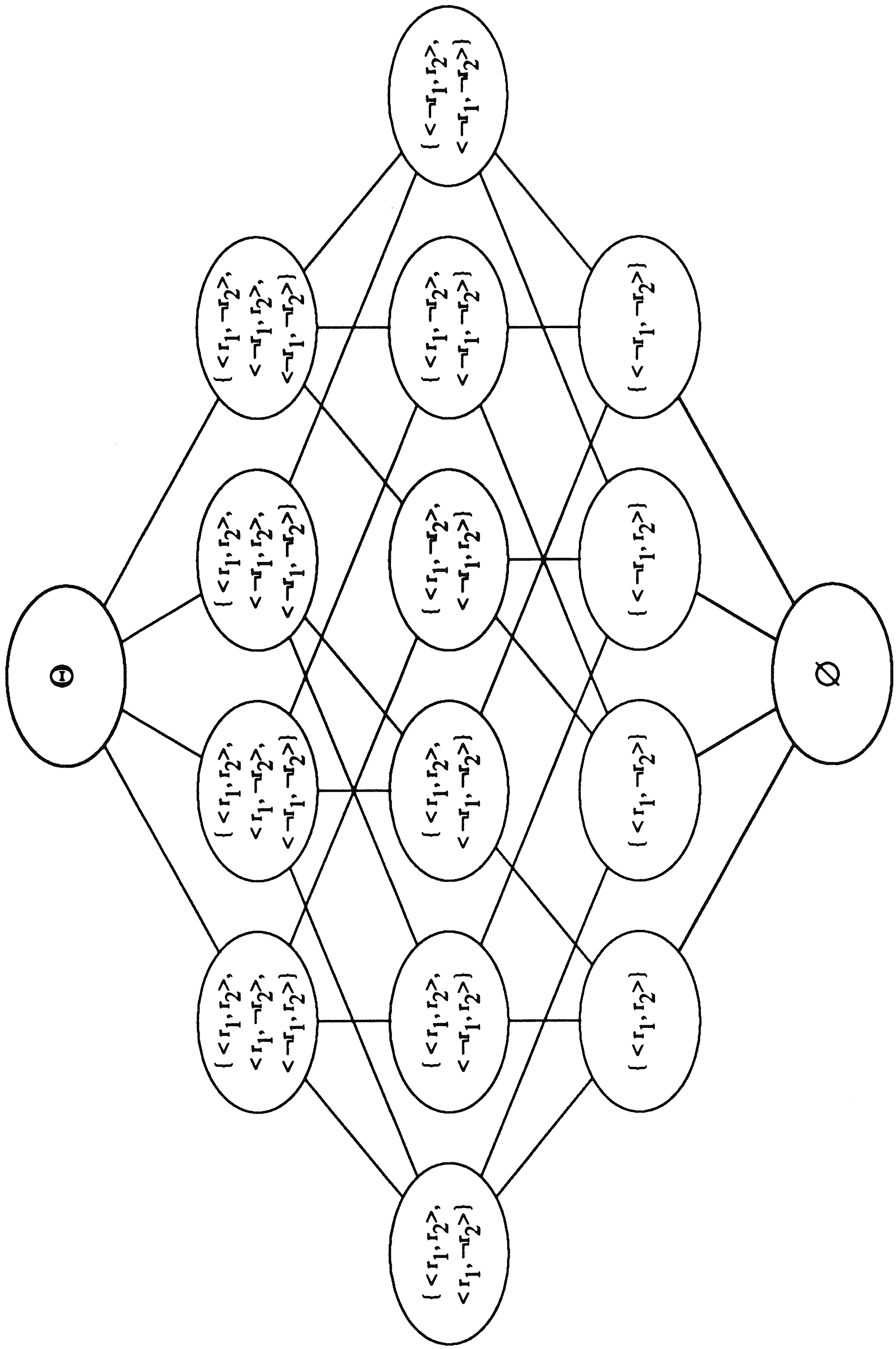

Figure 2. Hierarchical network of subsets for a two-vertex graph.

hierarchical network of subsets of the frame for a two-vertex graph. As mentioned above, we never have evidence supporting singletons and the subsets of the frame that are supported are not disjoint. In Figure 2 the last subset of the second row and the first two subsets of the third row are supported by one simple support function each. This is support offered against the belief in both vertices, i.e., support offered for the complement of the belief in both vertices, $< r_1, r_2 >^c$, support offered for the first vertex and support offered for the second vertex respectively. Because the supported subsets in the hierarchical network of our problem are not disjoint, we can not prune our network to a tree and use the scheme suggested by Gordon and Shortliffe.

The two other papers by Shafer and Logan [12] and Shafer, Shenoy, and Mellouli [13] concern the case of belief propagation in qualitative Markov trees only. Thus, the methods presented in these three papers are not applicable in the case with evidences in a non-prunable network of subsets.

Instead of propagating the belief in a hierarchical structure of subsets our algorithm reasons, separately for each instance of the frame, about the logical conditions of the completely specified path through the complete directed acyclic graph.

5. DEMPSTER'S RULE—THE CLASSIC ALGORITHM

Let us for convenience define the representation of a path as a conjunction of n propositions,

$$< x_1, x_2, \ldots, x_n > \triangleq x_1 \wedge x_2 \wedge \cdots \wedge x_n,$$

and define

$$\langle x_1, x_2, \ldots, x_n \rangle \triangleq x_1 \vee x_2 \vee \cdots \vee x_n$$

as a disjunction of n paths. We have

$$< x_1, x_2, \ldots, x_n > \wedge < y_1, y_2, \ldots, y_n > = < x_1 \wedge y_1, x_2 \wedge y_2, \ldots, x_n \wedge y_n > .$$

and

$$\begin{aligned}\langle x_1, x_2, \ldots, x_n \rangle \wedge \langle y_1, y_2, \ldots, y_n \rangle \\ = \langle x_1 \wedge y_1, x_1 \wedge y_2, \ldots, x_1 \wedge y_n, x_2 \wedge y_1, x_2 \wedge y_2, \ldots, x_2 \wedge y_n, \ldots, \\ x_n \wedge y_1, x_n \wedge y_2, \ldots, x_n \wedge y_n \rangle.\end{aligned}$$

In the problem of transfers between vertices in a graph, where a transfer might be possible only from a vertex with lower index to a vertex with higher index, our focus is on paths that may consist of several vertices. The frame of discernment is the set of all completely specified paths, $\Theta = \{< x_1, x_2, \ldots, x_n > | \forall i.\ x_i \in \{r_i, \neg r_i\}\}$, where r_i is the proposition of the evidence corresponding to vertex i in the graph, $x_i = r_i$ means that the vertex v_i is included in the path and $x_i = \neg r_i$ means that v_i is not included in the path.

We are interested in the problem where one begins with a basic probability assignment for those elements that belongs to the following subset of 2^{Θ}:

$$
\begin{aligned}
&\{< \theta_1, \ldots, \theta_{i-1}, r_i, \theta_{i+1}, \ldots, \theta_n >\} \\
&\quad \cup \Big\{ \langle < \theta_1, \ldots, \theta_{i-1}, \neg r_1, \theta_{i+1}, \ldots, \theta_n > , \\
&\qquad < \theta_1, \ldots, \theta_{j-1}, \neg r_j, \theta_{j+1}, \ldots, \theta_n > , \\
&\qquad < \theta_1, \ldots, \theta_{k-1}, r_k, \theta_{k+1}, \ldots, \theta_n > \rangle | \forall i, j, k.\, i < k < j \Big\},
\end{aligned}
$$

that is, we begin with positive evidence, e_i, for all vertices and negative evidence, e_{ij}, against all directed edges v_i to v_j where $i < j$—evidence that the path does not include v_i or that it does not include v_j or that it does include a vertex $v_k, i < k < j$, between v_i and v_j, thus excluding any direct transfer.

Thus, we have the following two types of evidences to consider.

1. The evidence e_i for every vertex in the graph. The bpa for the path with a single evidence e_i is

$$
m_i(< x_1, x_2, \ldots, x_n >) = \begin{cases} p_i, & (x_i = r_i) \wedge (\forall k | k \neq i.\ x_k = \theta_k) \\ 1 - p_i, & \forall k . x_k = \theta_k \\ 0, & \text{otherwise} \end{cases} ,
$$

2. The evidence e_{ij} against the edges between every two vertices in the graph. The corresponding bpa is here

$$
m_{ij}(< x_1, x_2, \ldots, x_n >) = \begin{cases} q_{ij}, & \big((x_i = r_i) \wedge (x_j = \neg r_j)\big) \vee \\ & \big((x_i = \neg r_i) \wedge (x_j = \theta_j)\big) \vee \\ & (\exists k | i < k < j.\ x_k = r_k) \\ 1 - q_{ij}, & \forall k.\ x_k = \theta_k \\ 0, & \text{otherwise} \end{cases} ,
$$

All these evidences are to be combined using Dempster's rule. The evidences can be combined in an arbitrary order because Dempster's rule is both associative and commutative.

5.1. Explaining the Classic Algorithm

We will seek the support and plausibility for all elements of Θ that are of the form $< x_1, x_2, \ldots, x_n >$ where $x_i \in \{r_i, \neg r_i\}$. For the sake of simplicity we shall first use Dempster's rule to separately fuse all positive evidences and all negative evidences,

$$\mathrm{Bel}_p = \oplus\{\mathrm{Bel}_{< \theta_1, \ldots, \theta_{i-1}, r_i, \theta_{i+1}, \ldots, \theta_n >}\},$$

$$\mathrm{Bel}_n = \oplus\Big\{\mathrm{Bel}_{\langle < \theta_1, \ldots, \theta_{i-1}, \neg r_i, \theta_{i+1}, \ldots, \theta_n >, < \theta_1, \ldots, \theta_{j-1}, \neg r_j, \theta_{j+1}, \ldots, \theta_n >, < \theta_1, \ldots, \theta_{k-1}, r_k, \theta_{k+1}, \ldots, \theta_n > \rangle} \Big| \forall i, j, k. i < k < j\Big\},$$

thus leaving the conflict creating fusion, $\mathrm{Bel}_p \oplus \mathrm{Bel}_n$, until last. The first of these fusions is shown in Figure 3 and Figure 4. The support and plausibility of all paths will then be calculated from the result of the last fusion.

In a fusion of two belief functions the representation in every intersection of focal elements is the conjunction of these focal elements' representations. The value of that intersection is the product of the values of the focal elements. In the upper left quadrant of Figure 4, for example, the result is derived from:

$$\begin{aligned}
&\langle < \neg r_1, \theta_2, \ldots, \theta_n >, < \theta_1, \neg r_2, \theta_3, \ldots, \theta_n > \rangle \\
&\wedge \langle < \neg r_1, \theta_2, \ldots, \theta_n >, < \theta_1, \theta_2, \neg r_3, \theta_4, \ldots, \theta_n >, \\
&\quad < \theta_1, r_2, \theta_3, \ldots, \theta_n > \rangle \\
&\quad = \langle < \neg r_1, \theta_2, \ldots, \theta_n >, < \neg r_1, \theta_2, \neg r_3, \theta_4, \ldots, \theta_n >, \\
&\qquad < \neg r_1, r_2, \theta_3, \ldots, \theta_n >, < \neg r_1, \neg r_2, \theta_3, \ldots, \theta_n >, \\
&\qquad < \theta_1, \neg r_2, \neg r_3, \theta_4, \ldots, \theta_n >, \varnothing \rangle \\
&\quad = \{\text{since the second, third and fourth} \\
&\qquad \text{elements are contained in the first}\} \\
&\quad = \langle < \neg r_1, \theta_2, \ldots, \theta_n >, < \theta_1, \neg r_2, \neg r_3, \theta_4, \ldots, \theta_n > \rangle
\end{aligned}$$

	$< \theta_1, r_2, \theta_3, \ldots, \theta_n >$ p_2	$< \theta_1, \ldots, \theta_n >$ $1 - p_2$
$< r_1, \theta_2, \ldots, \theta_n >$ p_1	$< r_1, r_2, \theta_3, \ldots, \theta_n >$ $p_1 p_2$	$< r_1, \theta_2, \ldots, \theta_n >$ $p_1(1 - p_2)$
$< \theta_1, \ldots, \theta_n >$ $1 - p_1$	$< \theta_1, r_2, \theta_3, \ldots, \theta_n >$ $(1 - p_1)p_2$	$< \theta_1, \ldots, \theta_n >$ $(1 - p_1)(1 - p_2)$

Figure 3. The first use of Dempster's rule on positive evidence.

	$\langle <\neg r_1, \theta_2, \ldots, \theta_n>,$ $<\theta_1, \theta_2, \neg r_3, \theta_4, \ldots, \theta_n>,$ $<\theta_1, r_2, \theta_3, \ldots, \theta_n>\rangle$ q_{13}	$<\theta_1, \ldots, \theta_n>$ $1 - q_{13}$
$\langle <\neg r_1, \theta_2, \ldots, \theta_n>,$ $<\theta_1, \neg r_2, \theta_3, \ldots, \theta_n>\rangle$ q_{12}	$\langle <\neg r_1, \theta_2, \ldots, \theta_n>,$ $<\theta_1, \neg r_2, \neg r_3, \theta_4, \ldots, \theta_n>\rangle$ $q_{12} \cdot q_{13}$	$\langle <\neg r_1, \theta_2, \ldots, \theta_n>$ $<\theta_1, \neg r_2, \theta_3, \ldots, \theta_n>\rangle$ $q_{12} \cdot (1 - q_{13})$
$<\theta_1, \ldots, \theta_n>$ $1 - q_{12}$	$\langle <\neg r_1, \theta_2, \ldots, \theta_n>,$ $<\theta_1, \theta_2, \neg r_3, \theta_4, \ldots, \theta_n>,$ $<\theta_1, r_2, \theta_3, \ldots, \theta_n>\rangle$ $(1 - q_{12}) \cdot q_{13}$	$<\theta_1, \ldots, \theta_n>$ $(1 - q_{12}) \cdot (1 - q_{13})$

Figure 4. The first use of Dempster's rule on negative evidence.

and

$$m(\langle < \neg r_1, \theta_2, \ldots, \theta_n > , < \theta_1, \neg r_2, \theta_3, \ldots, \theta_n > \rangle)$$

$$\cdot m(\langle < \neg r_1, \theta_2, \ldots, \theta_n > , < \theta_1, \theta_2, \neg r_3, \theta_4, \ldots, \theta_n > ,$$

$$< \theta_1, r_2, \theta_3, \ldots, \theta_n > \rangle) = q_{12} \cdot q_{13}.$$

The fusion, Figure 4, will result in a new basic probability assignment with basic probability numbers for all new representations. The basic probability number of $\langle < \neg r_1, \theta_2, \ldots, \theta_n > , < \theta_1, \neg r_2, \neg r_3, \theta_4, \ldots, \theta_n > \rangle$ for instance, is the normalized sum of values from all intersections with exactly this representation. In Figure 4 there are, of course, no other intersections with this representation and no conflict to cause a normalization. A new belief function is given by the new basic probability assignment and the belief of a proposition, A, is the sum of the basic probability numbers for that proposition, m(A), and all propositions that are proper subsets of A, m($B|B \subset A$). In our case, however, the situation is somewhat simpler because we are only seeking the support and plausibility of propositions that have no proper subsets.

Let us, for simplicity, observe the final fusion $\mathrm{Bel}_p \oplus \mathrm{Bel}_n$ in the case with three vertices, Figure 5. In each square the representation and its value is derived in the same way as above. The support and plausibility of all elements can be calculated as:

$$\forall x_i | x_i \in \{r_i, \neg r_i\}.\,\mathrm{Spt}(<x_1, x_2, \ldots, x_n>)$$

$$= \frac{1}{1-k} \sum_{A_i \cap A_j = <x_1, x_2, \ldots, x_n>} m(A_i) \cdot m(A_j),$$

$$\forall x_i | x_i \in \{r_i, \neg r_i\}.\,\mathrm{Pls}(<x_1, x_2, \ldots, x_n>)$$

$$= \frac{1}{1-k} \sum_{<x_1, x_2, \ldots, x_n> \in A_i \cap A_j} m(A_i) \cdot m(A_j)$$

where

$$k = \sum_{A_i \cap A_j = \varnothing} m(A_i) \cdot m(A_j),$$

$A_i, A_j \subset 2^{\Theta}$ are focal elements in the last fusion and

$$<x_1, x_2, \ldots, x_n> \in \langle y_1, y_2, \ldots, y_n \rangle \qquad \text{iff } \exists y_i | <x_1, x_2, \ldots, x_n> \in y_i,$$

$$<x_1, x_2, \ldots, x_n> \in <z_1, z_2, \ldots, z_n> \qquad \text{iff } \forall i.(z_i = x_i) \vee (z_i = \theta_i).$$

Due to the high computational complexity it is only possible to perform these computations for graphs consisting of very few vertices. This problem is solved by a new algorithm, where instead of performing all combinations step-by-step, the final result is derived directly by reasoning about the completely specified paths from the beginning.

5.2. An Example

Consider the path $<r_1, \neg r_2, r_3>$. Before we calculate the support and plausibility of the path we must calculate the conflict, k, in the final

	$<r_1, r_2, r_3>$ $p_1 \cdot p_2 \cdot p_3$	$<\theta_1, r_2, r_3>$ $(1 - p_1) \cdot p_2 \cdot p_3$	$<r_1, \theta_2, r_3>$ $p_1 \cdot (1 - p_2) \cdot p_3$	$<r_1, r_2, \theta_3>$ $p_1 \cdot p_2 \cdot (1 - p_3)$
$\langle <\neg r_1, \neg r_2, \theta_3>$ $<\neg r_1, \theta_2, \neg r_3>$ $<\theta_1, \neg r_2, \neg r_3> \rangle$ $q_{12} \cdot q_{13} \cdot q_{23}$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$\langle <\neg r_1, \neg r_2, \theta_3>$ $<\theta_1, \theta_2, \neg r_3> \rangle$ $(1 - q_{12}) \cdot q_{13} \cdot q_{23}$	$\emptyset$	$\emptyset$	$\emptyset$	$<r_1, r_2, \neg r_3>$
$\langle <\neg r_1, \theta_2, \neg r_3>$ $<\theta_1, \neg r_2, \theta_3> \rangle$ $q_{12} \cdot (1 - q_{13}) \cdot q_{23}$	$\emptyset$	$\emptyset$	$<r_1, \neg r_2, r_3>$	$\emptyset$
$\langle <\neg r_1, \theta_2, \theta_3>$ $<\theta_1, \neg r_2, \neg r_3> \rangle$ $q_{12} \cdot q_{13} \cdot (1 - q_{23})$	$\emptyset$	$<\neg r_1, r_2, r_3>$	$\emptyset$	$\emptyset$
$\langle <\theta_1, \neg r_2, \theta_3>$ $<\theta_1, \theta_2, \neg r_3> \rangle$ $(1 - q_{12}) \cdot (1 - q_{13}) \cdot q_{23}$	$\emptyset$	$\emptyset$	$<r_1, \neg r_2, r_3>$	$<r_1, r_2, \neg r_3>$
$\langle <\neg r_1, \theta_2, \theta_3>$ $<\theta_1, \theta_2, \neg r_3>$ $<\theta_1, r_2, \theta_3> \rangle$ $(1 - q_{12}) \cdot q_{13} \cdot (1 - q_{23})$	$<r_1, r_2, r_3>$	$<\theta_1, r_2, r_3>$	$<r_1, r_2, r_3>$	$<r_1, r_2, \theta_3>$
$\langle <\neg r_1, \theta_2, \theta_3>$ $<\theta_1, \neg r_2, \theta_3> \rangle$ $q_{12} \cdot (1 - q_{13}) \cdot (1 - q_{23})$	$\emptyset$	$<\neg r_1, r_2, r_3>$	$<r_1, \neg r_2, r_3>$	$\emptyset$
$<\theta_1, \theta_2, \theta_3>$ $(1 - q_{12}) \cdot (1 - q_{13}) \cdot (1 - q_{23})$	$<r_1, r_2, r_3>$	$<\theta_1, r_2, r_3>$	$<r_1, \theta_2, r_3>$	$<r_1, r_2, \theta_3>$

Figure 5. The last use of Dempster's rule: fusing positive and negative evidence.

fusion $\mathrm{Bel}_p \oplus \mathrm{Bel}_n$. The conflict is the sum of all contributions from all intersections $A_i \cap A_j = \emptyset$, Figure 5;

$$k = \cdots = p_1 \cdot p_2 \cdot q_{12} + p_1 \cdot (1 - p_2) \cdot p_3$$
$$\cdot (q_{12} \cdot q_{13} + q_{13} \cdot q_{23} - q_{12} \cdot q_{13} \cdot q_{23})$$
$$+ p_2 \cdot p_3 \cdot q_{23} - p_1 \cdot p_2 \cdot p_3 \cdot q_{12} \cdot q_{23}.$$

$\langle \theta_1, \theta_2, r_3 \rangle$ $(1-p_1)\cdot(1-p_2)\cdot p_3$	$\langle \theta_1, r_2, \theta_3 \rangle$ $(1-p_1)\cdot p_2\cdot(1-p_3)$	$\langle r_1, \theta_2, \theta_3 \rangle$ $p_1\cdot(1-p_2)\cdot(1-p_3)$	$\langle \theta_1, \theta_2, \theta_3 \rangle$ $(1-p_1)\cdot(1-p_2)\cdot(1-p_3)$
$\langle \neg r_1, \neg r_2, r_3 \rangle$	$\langle \neg r_1, r_2, \neg r_3 \rangle$	$\langle r_1, \neg r_2, \neg r_3 \rangle$	$(\langle \neg r_1, \neg r_2, \theta_3 \rangle$ $\langle \neg r_1, \theta_2, \neg r_3 \rangle$ $\langle \theta_1, \neg r_2, \neg r_3 \rangle)$
$\langle \neg r_1, \neg r_2, r_3 \rangle$	$\langle \theta_1, r_2, \neg r_3 \rangle$	$\langle r_1, \theta_2, \neg r_3 \rangle$	$(\langle \neg r_1, \neg r_2, \theta_3 \rangle$ $\langle \theta_1, \theta_2, \neg r_3 \rangle)$
$\langle \theta_1, \neg r_2, r_3 \rangle$	$\langle \neg r_1, r_2, \neg r_3 \rangle$	$\langle r_1, \neg r_2, \theta_3 \rangle$	$(\langle \neg r_1, \theta_2, \neg r_3 \rangle$ $\langle \theta_1, \neg r_2, \theta_3 \rangle)$
$\langle \neg r_1, \theta_2, r_3 \rangle$	$\langle \neg r_1, r_2, \theta_3 \rangle$	$\langle r_1, \neg r_2, \neg r_3 \rangle$	$(\langle \neg r_1, \theta_2, \theta_3 \rangle$ $\langle \theta_1, \neg r_2, \neg r_3 \rangle)$
$\langle \theta_1, \neg r_2, r_3 \rangle$	$\langle \theta_1, r_2, \neg r_3 \rangle$	$(\langle r_1, \neg r_2, \theta_3 \rangle$ $\langle r_1, \theta_2, \neg r_3 \rangle)$	$(\langle \theta_1, \neg r_2, \theta_3 \rangle$ $\langle \theta_1, \theta_2, \neg r_3 \rangle)$
$(\langle \neg r_1, \theta_2, r_3 \rangle$ $\langle \theta_1, r_2, r_3 \rangle)$	$\langle \theta_1, r_2, \theta_3 \rangle$	$(\langle r_1, \theta_2, \neg r_3 \rangle$ $\langle r_1, r_2, \theta_3 \rangle)$	$(\langle \neg r_1, \theta_2, \theta_3 \rangle$ $\langle \theta_1, \theta_2, \neg r_3 \rangle$ $\langle \theta_1, r_2, \theta_3 \rangle)$
$(\langle \neg r_1, \theta_2, r_3 \rangle$ $\langle \theta_1, \neg r_2, r_3 \rangle)$	$\langle \neg r_1, r_2, \theta_3 \rangle$	$\langle r_1, \neg r_2, \theta_3 \rangle$	$(\langle \neg r_1, \theta_2, \theta_3 \rangle$ $\langle \theta_1, \neg r_2, \theta_3 \rangle)$
$\langle \theta_1, \theta_2, r_3 \rangle$	$\langle \theta_1, r_2, \theta_3 \rangle$	$\langle r_1, \theta_2, \theta_3 \rangle$	$\langle \theta_1, \theta_2, \theta_3 \rangle$

Figure 5. *Continued.*

The support is calculated as the normalized sum of all contributions from the intersection whose representation is identical with the path. Thus the support of $< r_1, \neg r_2, r_3 >$ is the contributions, in Figure 5, from row 3 column 3, row 5 column 3, and row 7 column 3;

$$\mathrm{Spt}(< r_1, \neg r_2, r_3 >) = \frac{1}{1-k}(p_1 \cdot (1-p_2) \cdot p_3 \cdot q_{12} \cdot (1-q_{13}) \cdot q_{23}$$
$$+p_1 \cdot (1-p_2) \cdot p_3 \cdot (1-q_{12}) \cdot (1-q_{13}) \cdot q_{23}$$

$$
\begin{aligned}
&+p_1 \cdot (1 - p_2) \cdot p_3 \cdot q_{12} \cdot (1 - q_{13}) \cdot (1 - q_{23})) \\
&= \frac{1}{1-k}(p_1 \cdot (1 - p_2) \cdot p_3 \cdot (1 - q_{13}) \\
&\cdot (q_{12} + (1 - q_{12}) \cdot q_{23})).
\end{aligned}
$$

When calculating the plausibility we normalize the sum of all contributions from the intersections in which representations the path is contained. These are the 16 intersections of rows 3, 5, 7, 8 and columns 3, 5, 7, 8, Figure 5.

Take for instance the intersection in row 5 column 7:

$$< r_1, \neg r_2, r_3 > \in \langle < r_1, \neg r_2, \theta_3 > , < r_1, \theta_2, \neg r_3 > \rangle$$

since

$$< r_1, \neg r_2, r_3 > \in y_1 \quad (= < r_1, \neg r_2, \theta_3 >)$$

which is true since

$$
\begin{cases}
z_1 = x_1 \ (= r_1) \\
z_2 = x_2 \ (= \neg r_2) \\
z_3 = \theta_3 .
\end{cases}
$$

Thus $< r_1, \neg r_2, r_3 >$ is contained in $\langle < r_1, \neg r_2, \theta_3 > , < r_1, \theta_2, \neg r_3 > \rangle$ and the value of the intersection in row 5 column 7 is contributing to the plausibility of $< r_1, \neg r_2, r_3 >$. The plausibility becomes, after some simplification:

$$\text{Pls}(< r_1, \neg r_2, r_3 >) = \cdots = \frac{1}{1-k}(1 - p_2) \cdot (1 - q_{13}).$$

6. DEMPSTER'S RULE—THE NEW ALGORITHM

We are now ready to give an intuitive presentation of our algorithm for calculating the support and plausibility for all elements, A, of 2^Θ that are of the form $< x_1, x_2, \ldots, x_n >$ where $x_i \in \{r_i, \neg r_i\}$, i.e., the completely specified paths of Θ. The new algorithm is built on an expression for the final result of support and plausibility, i.e., we only have to evaluate this expression instead of all stepwise combinations. The algorithm, when used symbolically, calculate the symbolic structure of the support and plausibility for a path derived through summation of intersections in the final fusion, $\text{Bel}_p \oplus \text{Bel}_n$, of the classic algorithm. This means that the new algorithm can comparatively quickly calculate the answers which had to be calculated through a lot of time-consuming fusions and pattern-matching summations in the classic algorithm. We will first explain the mathematical reasoning behind this algorithm, which calculates support and plausibility in the following steps: unnormalized plausibility, unnormalized support, conflict, and finally plausibility and support normalized by the conflict.

6.1. Plausibility

Let us start with the plausibility and see what is sufficient to make a path plausible. Plausibility for a path means to which degree this path is possible, i.e., to which degree no known factors speak against this path. There are only two types of items which speak against a path—the positive evidence for vertices that are not included in the path and the negative evidence against edges between vertices that are included in the path. This means that the degree to which we do not assign support to these evidences equals the degree to which the path is possible. The algorithm for plausibility is then

$$\mathrm{Pls}(S) = \mathrm{Pls}^*(S)/(1-k)$$

where k is the conflict and $\mathrm{Pls}^*(\cdot)$ is the unnormalized plausibility

$$\mathrm{Pls}^*(S) = \prod_{\forall i \mid v_i \notin S} (1-p_i) \cdot \prod_{\forall i} (1 - q_{s_i, s_{i+1}}),$$

where q_{ij} is the degree of doubt of the edge between vertices v_i and v_j and v_{s_i} is the i:th vertex in the path S.

6.2. Support

The algorithm for support is much more complicated than the one for plausibility. It is not only necessary to find out which evidence speaks against the path; it is also necessary to insist that the evidence of the vertices and edges that are included in the path speaks in favor of it.

While each of the evidences supports only one proper subset of Θ, i.e., corresponding belief functions are simple support functions, we will say for the sake of simplicity that the evidence e_i is false (true) when we mean that the proposition according to the proper subset is false (true). The same holds for the evidences e_{ij}.

6.2.1. EXPLAINING THE ALGORITHM FOR SUPPORT Assuming the path includes m vertices, we first realize that the following three statements have to be true:

1. Every vertex in the path has to be visited.
   - (a) For the first and the last vertices in the path we only have one possibility: the evidences e_{s_1} and e_{s_m} are true and the support for this is $p_{s_1} \cdot p_{s_m}$, for $m > 1$ and p_{s_1}, $m = 1$.

(b) For every intermediate vertex v_{s_i} in S there are two different possibilities:
   (i) The evidence e_{s_i} is true. The support for this is p_{s_i}.
   (ii) The evidence e_{s_i} may be false, but the evidence against edges are speaking against all other ways from the last vertex visited before v_{s_i}, to the first vertex visited after v_{s_i}. The possibility that e_{s_i} may be false is $(1 - p_{s_i})$.

2. The transitions between consecutive vertices in the path are possible, i.e., the evidence against those edges has to be false. The possibility for this is

$$\prod_{i=1}^{m-1} (1 - q_{s_i, s_{i+1}}).$$

3. No vertex outside the path is permitted to be visited. We first state that the evidences e_i for vertices outside the path have to be false. The possibility for this is

$$\prod_{\forall i | v_i \notin S} (1 - p_i).$$

But even if these evidences may be false, we can not be sure that a vertex outside S is not visited. In order to guarantee this we also make the following three statements:

(a) No transition is possible from vertices before v_{s_1} to this vertex, i.e., all evidences against edges from vertices before v_{s_1} to v_{s_1} are true. The support for this is:

$$\prod_{i<s_1} q_{i, s_1}.$$

This statement assures that we enter the path at v_{s_1}.

(b) No transition is possible from v_{s_m} to vertices after this vertex. This is to assure that v_{s_m} is the last vertex in the path. The support for this is:

$$\prod_{i>s_m} q_{s_m, i}.$$

(c) For the vertices not belonging to S that are located between vertices in S we state that no transition is possible to these vertices from vertices in S, or if such a transition is possible then it is not possible to join the next vertex in the path we have stated to be true.

The support is now received by multiplying the contributions from the three statements. Let us illustrate this with an example.

Let the graph G consist of five vertices $v_1, \ldots, v_5$. We shall compute the support and plausibility for the path $< r_1, \neg r_2, r_3, r_4, \neg r_5 >$. The unnormalized plausibility is easily derived in the way described above:

$$\text{Pls}^*(< r_1, \neg r_2, r_3, r_4, \neg r_5 >) = (1-p_2)\cdot(1-p_5)\cdot(1-q_{13})\cdot(1-q_{34}).$$

When computing the support we apply the three statements above. From statement (1a) we get the factor $p_1 \cdot p_4$. Considering statement (1b) forces us to break down the calculations into two parts:

(i) We state e_3 to be true.

(ii) We do not state e_3 to be true.

The factor calculated in (3c), in order to prevent visiting a vertex between the first and last vertices of the path which does not belong to the path, will differ depending on which vertices we have stated to be true, therefore we calculate the factors from the other statements separately for (i) and (ii) and then sum up the two contributions.

We begin with (i).

When e_3 is true, the factor from this statement is p_3. From (2) we get the factor $(1 - q_{13}) \cdot (1 - q_{34})$. Statement (3) states that the evidences e_2 and e_5 have to be false, giving us the factor $(1 - p_2) \cdot (1 - p_5)$. Statement (3a) can be disregarded while the first vertex in the path is the first vertex in the graph and from (3b) we get the factor q_{45}.

Let us now regard (3c) which states either that a transition from v_1 to v_2 is not allowed, which gives us the factor q_{12}, or that if a transition from v_1 to v_2 is allowed, then it must be impossible to reach the next vertex in the path stated to be true, which according to our assumption (1b) is v_3. This gives us the factor $(1 - q_{12}) \cdot q_{23}$, i.e., the total factor from (3c) is $q_{12} + (1 - q_{12}) \cdot q_{23}$. We have now calculated the first term of the support $p_1 \cdot p_4 \cdot p_3 \cdot (1 - q_{13}) \cdot (1 - q_{34}) \cdot (1 - p_2) \cdot (1 - p_5) \cdot q_{45} \cdot (q_{12} + (1 - q_{12}) \cdot q_{23})$.

Let us calculate the second term, (ii), where we do not state e_3 to be true.

The possibility for this is $(1 - p_3)$. The factors (1a), (2), and (3a–b) in this term are the same as in the term above and (3c) is in this case implied in (1b), hence it is enough to calculate (1b). We have the following two possibilities:

(1) transition from v_1 to v_2 or v_4 is impossible, which implies that the only path from v_1 to v_4 is $v_1 - v_3 - v_4$. This gives us the factor $q_{12} \cdot q_{14}$.

(2) transition from v_1 to v_4 is impossible but we allow a transition from v_1 to v_2 but not from v_2 to v_3 or v_4, giving us the factor $(1 - q_{12}) \cdot q_{14} \cdot q_{23} \cdot q_{24}$.

The second term for the support is then

$$p_1 \cdot p_4 \cdot (1 - p_3) \cdot (1 - q_{13}) \cdot (1 - q_{34}) \cdot q_{45} \cdot (1 - p_2) \cdot (1 - p_5)$$
$$\cdot (q_{12} \cdot q_{14} + (1 - q_{12}) \cdot q_{14} \cdot q_{23} \cdot q_{24})$$

and we end up with the unnormalized support

$$\begin{aligned}&\mathrm{Sup}^*(<r_1, \neg r_2, r_3, r_4, \neg r_5>)\\&\quad = p_1 \cdot (1-p_2) \cdot p_4 \cdot (1-p_5) \cdot (1-q_{13})\\&\qquad \cdot (1-q_{34}) \cdot q_{45} \cdot (p_3 \cdot (q_{12} + (1-q_{12}) \cdot q_{23}) + (1-p_3) \cdot q_{14}\\&\qquad \cdot (q_{12} + (1-q_{12}) \cdot q_{23} \cdot q_{24})).\end{aligned}$$

The normalized support becomes

$$\mathrm{Sup}(<r_1, \neg r_2, r_3, r_4, \neg r_5>) = \mathrm{Sup}^*(<r_1, \neg r_2, r_3, r_4, \neg r_5>)/(1-k)$$

where k is the conflict.

In Section 6.2.2 we present a detailed analysis of the algorithm for support, followed in Section 6.2.3 by the algorithm itself. The reader may skip these sections on a first reading and continue with Section 6.3 on conflict.

6.2.2. A DETAILED ANALYSIS OF THE ALGORITHM FOR SUPPORT First some useful definitions:

$$\begin{aligned}m^-(\omega, i) &\triangleq \min(i) | \omega_i = r_i, & 1 \le i \le n,\\ m^-(\omega, i, j) &\triangleq \min(i) | \omega_i = r_i, & 1 \le i, j \le n, i > j\\ m^+(\omega, i) &\triangleq \max(i) | \omega_i = r_i, & 1 \le i \le n,\\ m^+(\omega, i, j) &\triangleq \max(i) | \omega_i = r_i, & 1 \le i, j \le n, i < j.\end{aligned}$$

Thus, $m^-(x, i)$ is the first vertex in the path, v_{s_1}, and $m^-(x, i, j)$ is the first vertex in the path of those with index larger than j.

The algorithm can be broken down into three different parts.

For the first part we have the same argument as with the plausibility, i.e., the evidence which speaks against the path must be false, thus the same terms as in the plausibility.

The second part of the algorithm is to assure that the first and last vertices of the path actually are the first and last vertices included in the path, i.e., that there is evidence against edges to the first vertex of the path from any vertices in the graph before the path's first vertex, that the path's first and last vertices are included in the path, and that there is evidence against edges from the last vertex of the path to any vertices in the graph after the path's last vertex. This gives us the terms:

$$\left(\prod_{\forall j | 1 \le j < m^-(x,i)} q_{j, m^-(x,i)} \right) \cdot p_{m^-(x,i)}$$

and

$$\begin{cases} p_{m^+(x,i)}, & m^-(x,i) \ne m^+(x,i) \\ 1, & m^-(x,i) = m^+(x,i) \end{cases} \cdot \left(\prod_{\forall j | m^+(x,i) < j \le n} q_{m^+(x,i), j} \right).$$

The third part of the algorithm concerns the transfers from the first vertex of the path until the last one. The positive evidence of every internal vertex of the path, i.e., the vertices $x_j = r_j$ where $m^-(x,i) < j < m^+(x,i)$, is to some degree committed in favor of the path and is for the remainder uncommitted. However, for each combination of statements for the internal vertices, i.e., internal vertices stated or not stated to be true for the combination, we have support for the path given the right conditions for the edges. Thus, we have to sum up the contribution from all the combinations;

$$\forall\left(\bigwedge_{\forall j \left| \begin{smallmatrix} m^-(x,i)<j<m^+(x,i) \\ x_j=r_j \end{smallmatrix} \right.} y_j\right)\Bigg| y_j = r_j, \neg r_j$$

where

$$\bigwedge_{\forall j \left| \begin{smallmatrix} m^-(x,i)<j<m^+(x,i) \\ x_j=r_j \end{smallmatrix} \right.} y_j$$

is a general description of a combination of statements. As an example, consider the path $< r_1, r_2, \neg r_3, r_4, r_5 >$. We have $m^-(x,i) = 1$, $m^+(x,i) = 5$ and $x_j = r_j, j \neq 3$. The general description of a combination is $(y_2 \wedge y_4)$ where

$$\forall(y_2 \wedge y_4)|y_j = r_j, \neg r_j$$

yields the set of all combinations, $\{r_2 \wedge r_4, r_2 \wedge \neg r_4, \neg r_2 \wedge r_4, \neg r_2 \wedge \neg r_4\}$.

The contribution from each combination depends on the positive evidence for that combination, the term

$$\prod_{\forall k \left| \begin{smallmatrix} m^-(x,i)<k<m^+(x,i) \\ x_k=r_k \end{smallmatrix} \right.} \begin{cases} p_k, y_k = r_k \\ 1 - p_k, y_k = \neg r_k \end{cases},$$

and the negative evidence given by the following necessary conditions for that combination.

The first condition is that all internal vertices must be visited. Hence, for each sequence of vertices among the internal vertices, that are not stated to be true in this combination, Figure 6, we must block all forbidden edges. These are edges from a vertex v_i to a vertex v_j where v_i is in the sequence or the last vertex before the sequence, v_j is in the sequence or the first vertex after the sequence and where there is a vertex v_k such that v_k is in

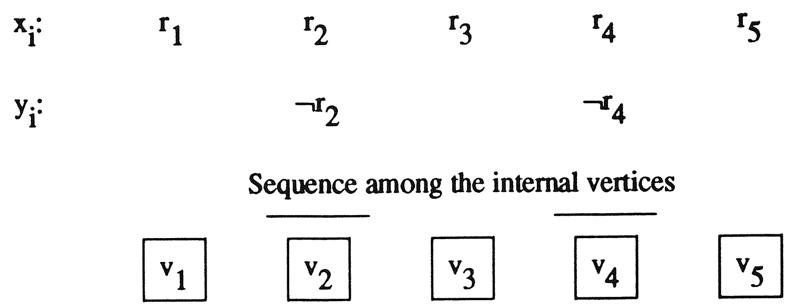

Figure 6. Vertices v_2 and v_4 form a sequence because v_3 is not in the path.

the sequence, i.e., the sequence of internal vertices not stated to be true in the present combination, and $i < k < j$. It is accomplished by the term

$$\left(\prod_{\forall k \,\left|\, \substack{m^-(x,i)<k<m^+(x,i) \\ x_k = r_k \\ y_k = \neg r_k}\right.} \left(\prod_{\forall m \,\left|\, \substack{k<m\le \min(m^-(y,i,k),\, m^+(x,i)) \\ x_m = r_m}\right.} q_{m^+(x,i,k),m} \right) \right)$$

where $m^+(x, i, k)$ is the last vertex in the path before the vertex v_k not stated to be true and $\min(m^-(y, i, k), m^+(x, i))$ is the first vertex after the sequence of vertices not stated to be true. As an example of the first condition, consider again the path $< r_1, r_2, \neg r_3, r_4, r_5 >$ now for the combination $(y_2 \wedge y_4) = (\neg r_2 \wedge \neg r_4)$. In Figure 7 the necessary edges are blocked. These are the edges v_1 to v_4, v_1 to v_5, and v_2 to v_5. Vertices v_1 and v_2 are in or the last vertex before the sequence, v_4 and v_5 are in or the first vertex after the sequence, and there is at least one vertex between the two vertices of the edge, in these cases v_2, v_2 and v_4, and vertex v_4 respectively.

The second and final condition will assure that, between the first and last vertex of the path, no vertices other than those in the path are visited. The evidence against every edge from a vertex v_i to a vertex v_j where v_i is included in the path, v_j, $i < j < s_n$, is not included in the path and where there are no internal vertices v_k, $i < k < j$, that are stated to be true to the path in the present combination, Figure 8, is to some degree committed in favor of the path and is for the remainder uncommitted. However, for each combination of statements for the internal vertices we will have support for the path from all combinations of statements for sets of edges from earlier vertices to an internal vertex, Figure 9, given correct condi-

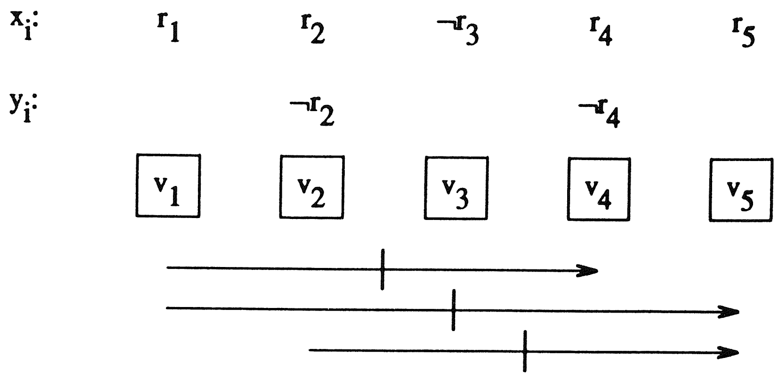

Figure 7. Example of the first condition for the combination $(\neg r_2 \wedge \neg r_4)$.

tions for the edges from this internal vertex. The evidence against set of edges, from vertices v_i to a vertex v_j where there for each v_i are no internal vertices v_k, $i < k < j$, stated to be true, is considered to be true if all edges in the set are blocked. Hence, for each combination of statements for internal vertices we will sum up the contribution from all

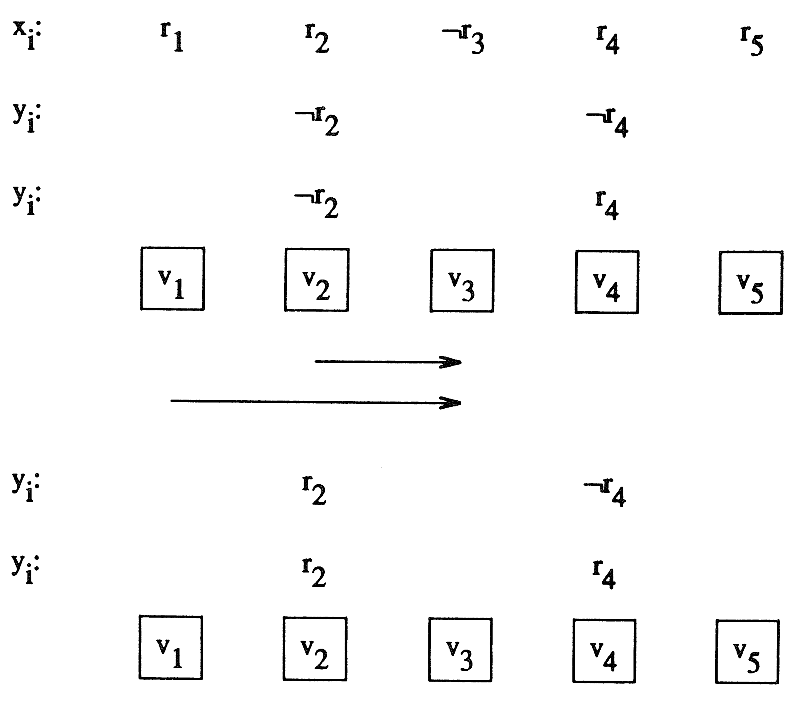

Figure 8. Possible edges to v_3 for different combinations.

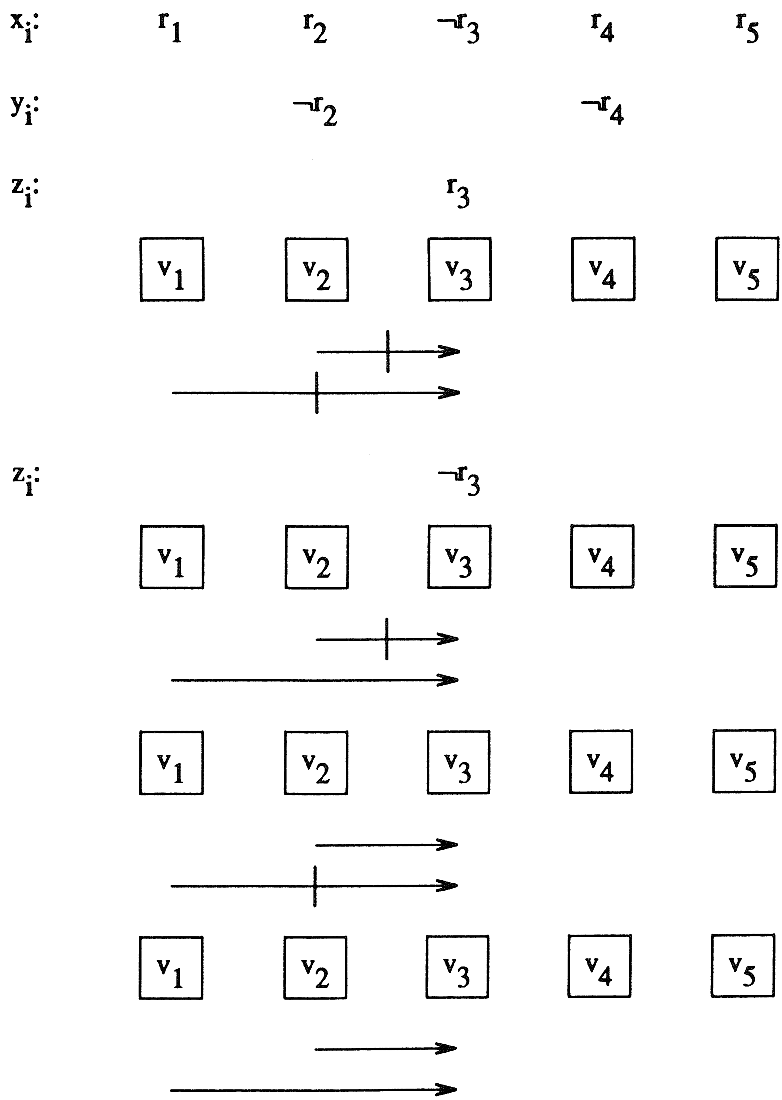

Figure 9. Two different states, z_3 ($= r_3, \neg r_3$), with one and three alternatives.

combinations of statements for the set of edges to these vertices. If the evidence against the set of edges to one of these vertices is not stated to be true, then we should sum up the contribution from all those alternatives of the edges to that vertex where at least the evidence of one of the edges is not stated to be true. An example of when the evidence against the set of edges to one vertex is not stated to be true, i.e., when all edges in the set are not blocked, is the three alternatives of the second combination of set of edges to vertex 3 in Figure 9. We must also take into account the necessary conditions on the edges from that vertex. However, because the conditions are the same for all alternatives when at least one of the edges to the vertex is not blocked, as with the three alternatives for the second combination in Figure 9, we are able to view all these alternatives in a set of edges not stated to be true as one generalized edge to the vertex that is not stated to be true for the present combination of internal vertices. The necessary conditions are that the edges from the vertex to all internal vertices not stated to be true in a subsequent sequence and to the first vertex after the sequence are blocked (Figure 10). Its contribution is

$$\left(1 - \prod_{\forall u \,\middle|\, \substack{\max(m^+(y,i,t),\, m^-(x,i)) \le u < t \\ (z_u = \neg r_u) \vee (x_u = r_u)}} q_{u,t}\right),$$

where t is the index of the vertex not included in the path, $u < t < n_s$, z_u marks whether or not all edges from vertices v_j in the path to vertex v_u, $j < u < s_n$, where there are no internal vertices v_k, $j < k < u$, stated to be true, are blocked. The necessary condition on the edges from vertex v_t are:

$$\prod_{\forall v \,\middle|\, \substack{t < v \le \min(m^-(y,i,t),\, m^+(x,i)). \\ x_v = r_v}} q_{t,v}.$$

The final alternative that all edges to the vertex are blocked, as in the first combination of set of edges in Figure 9, involve no conditions. Its contribu-

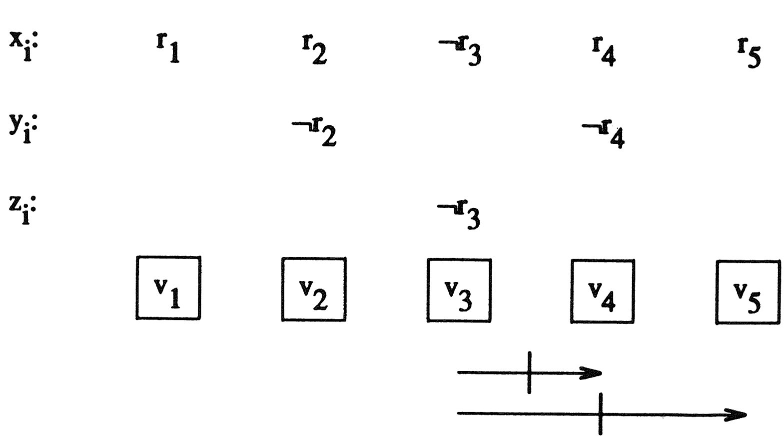

Figure 10. Conditions on the edges from vertex 3.

tion is:

$$\prod_{\forall u \left| \substack{\max(m^+(y,i,t),\, m^-(x,i)) \le u < t \\ (z_u = \neg r_u) \vee (x_u = r_u)} \right.} q_{u,t}.$$

As an example of the second condition, consider the path $< r_1, \neg r_2, \neg r_3, r_4, r_5 >$ for the combination of statements for the internal vertex $y_4 = \neg r_4$ and the combination of statements for set of edges $(z_2 \wedge z_3) = (\neg r_2 \wedge \neg r_3)$, Figure 11. That is, nothing speaks in favor of vertex 4. Furthermore, consider the edges where there are no internal vertices, between the vertices of the edge, stated to be true in the present combination. There is nothing that speaks against that there is at least one of these edges from a vertex in the path to vertices 2 and 3 respectively that is not blocked. If there is an edge to vertex 2 then it must be coming from vertex 1. The necessary condition is that all edges from vertex 2 to all internal

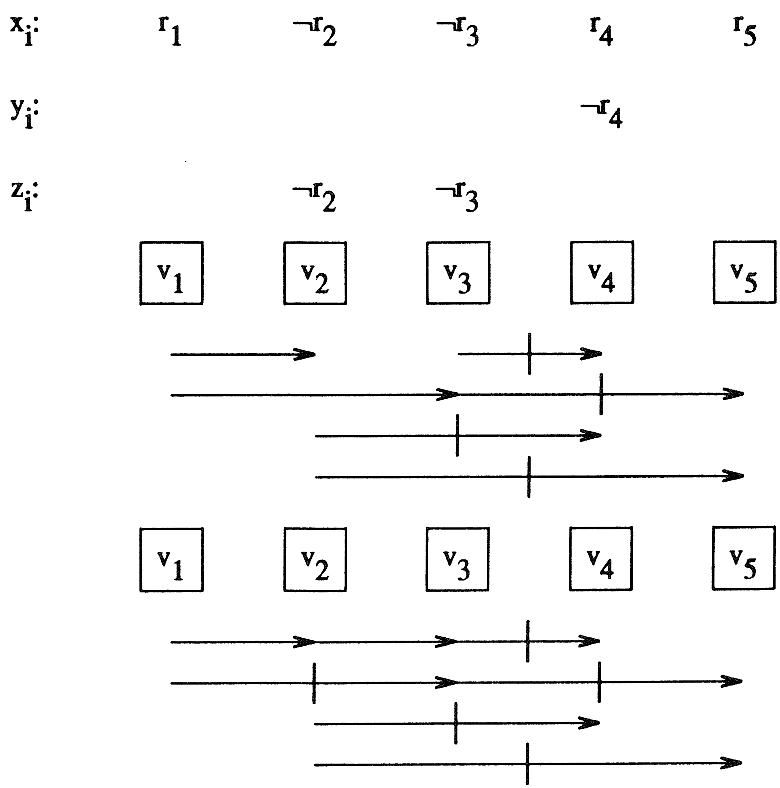

Figure 11. The second condition for the combination $y_4 = \neg r_4, (z_2 \wedge z_3) = (\neg r_2 \wedge \neg r_3)$.

vertices not stated to be true in a subsequent sequence and the first vertex after the sequence are blocked. Because vertex 4 is not stated to be true in this combination it is necessary to block the edges v_2 to v_4 and v_2 to v_5. There are two different edges to vertex 3, v_1 to v_3 and v_2 to v_3. At least one of these two should not be blocked. This gives us three different alternatives, v_1 to v_3 and not v_2 to v_3, not v_1 to v_3 and v_2 to v_3 and finally v_1 to v_3 and v_2 to v_3. The corresponding term becomes $(1 - q_{13}) \cdot q_{23} + q_{13} \cdot (1 - q_{23}) + (1 - q_{13}) \cdot (1 - q_{23})$, rewritten as $(1 - q_{13}) + q_{13} \cdot (1 - q_{23})$ it is understood as v_1 to v_3 or if not v_1 to v_3 then v_2 to v_3, as described in Figure 11. Rewriting the term as $1 - q_{13} \cdot q_{23}$ can be interpreted as one generalized edge to vertex 3 whose evidence is not stated to be true. This is the way it is rewritten in the algorithm. The condition for vertex 3 is of the same type as for vertex 2, here that the edges v_3 to v_4 and v_3 to v_5 are blocked.

6.2.3. THE ALGORITHM FOR SUPPORT The algorithm for support can then be summarized as

$$\forall x_i | x_i \in \{r_i, \neg r_i\}. \mathrm{Spt}(< x_1, x_2, \ldots, x_n >)$$

$$= \frac{1}{1 - k_n} \cdot p_{m^-(x,i)} \cdot \begin{cases} p_{m^+(x,i)}, & m^-(x,i) \neq m^+(x,i) \\ 1, & m^-(x,i) = m^+(x,i) \end{cases}$$

$$\cdot \left(\prod_{\forall i | x_i = \neg r_i} (1 - p_i) \right) \cdot \left(\prod_{\forall j | 1 \leq j < m^-(x,i)} q_{j,m^-(x,i)} \right)$$

$$\cdot \left(\prod_{\forall j | m^+(x,i) < j \leq n} q_{m^+(x,i),j} \right) \cdot \left(\prod_{\forall j \left| \substack{j \neq m^+(x,i) \\ x_j = r_j} \right.} 1 - q_{j,m^-(x,i,j)} \right)$$

$$\cdot \left(\sum_{\forall \left(\bigwedge_{\forall j \left| \substack{m^-(x,i) < j < m^+(x,i) \\ x_j = r_j} \right.} y_j \right) \middle| y_j = r_j, \neg r_j} \left(\prod_{\forall k \left| \substack{m^-(x,i) < k < m^+(x,i) \\ x_k = r_k} \right.} \begin{cases} p_k, & y_k = r_k \\ 1 - p_k, & y_k = \neg r_k \end{cases} \right) \right.$$

$$\cdot \left(\prod_{\forall k \left| \substack{m^-(x,i)<k<m^+(x,i) \\ x_k = r_k \\ y_k = \neg r_k} \right.} \left(\prod_{\forall m \left| \substack{k<m\le \min(m^-(y,i,k),\, m^+(x,i)) \\ x_m = r_m} \right.} q_{m^+(x,i,k),m} \right) \right)$$

$$\cdot \left(\sum_{\forall \left(\bigwedge_{\forall s \left| \substack{m^-(x,i)<s<m^+(x,i) \\ x_s = \neg r_s} \right.} z_S \right) \left| z_S = r_S, \neg r_S \right.} \left(\prod_{\forall t \left| \substack{m^-(x,i)<t<m^+(x,i) \\ x_t = \neg r_t} \right.} \begin{cases} \xi, & z_t = r_t \\ \psi, & z_t = \neg r_t \end{cases} \right) \right)$$

where

$$\xi = \prod_{\forall u \left| \substack{\max(m^+(y,i,t),\, m^-(x,i)) \le u < t \\ (z_u = \neg r_u) \vee (x_u = r_u)} \right.} q_{u,t}$$

and

$$\psi = \left(1 - \prod_{\forall u \left| \substack{\max(m^+(y,i,t),\, m^-(x,i)) \le u < t \\ (z_u = \neg r_u) \vee (x_u = r_u)} \right.} q_{u,t} \right)$$

$$\cdot \prod_{\forall v \left| \substack{t<v\le \min(m^-(y,i,t),\, m^+(x,i)) \\ x_v = r_v} \right.} q_{t,v}$$

and k_n is the conflict in the n-vertex graph.

6.3. Conflict

The conflict indicates the amount of the total mass consisting of contradictory evidences, i.e., evidences whose intersection is the empty set, $\varnothing$. This means that we actually compute the support for $\varnothing$, but as we do not want to assign any belief to an impossible event, this is denoted conflict,

$$\text{Conf} = \sum_{A=\varnothing} m(A).$$

6.3.1. EXPLAINING THE ALGORITHM FOR CONFLICT In our case the conflict arises when combining the evidences e_{ij} with the evidences concerning vertices $v_i, \ldots, v_j$. The calculations are based on the formula:

$$\text{Conf}(\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_{n+1}) = \text{Conf}(\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n) + \text{Conf}(\tilde{e}_1 \oplus \tilde{e}_2 \oplus \ldots \oplus \tilde{e}_n, \tilde{e}_{n+1}),$$

where $\tilde{e}_i$ are arbitrary evidences and $\tilde{e}_i \oplus \tilde{e}_j$ the combined evidence from $\tilde{e}_i$ and $\tilde{e}_j$. We here in fact mean the basic probability assignment for the evidences, but for simplicity we use the denotation for evidence. The formula above means that when we add new evidences to already combined evidences the new conflict is obtained as the sum of the earlier conflict and a contribution from the new evidences. The conflict can never decrease when bringing in new evidences. For the sake of clarity we assume that the combination of evidences take place stepwise in the following order:

$$e_1 \oplus e_2 \oplus e_{12} \oplus e_3 \oplus e_{23} \oplus e_{13} \oplus e_4 \oplus \ldots \oplus e_n \oplus e_{n-1n} \oplus \ldots \oplus e_{1n}.$$

The positive evidences e_i are brought into the combination in increasing order of i, but between the e_i all negative evidences e_{ij} are regarded in such a way that e_k is followed by all e_{ik} where $i < k$. This means that the e_i never give rise to any conflict when they are brought into the combination which on the other hand the e_{ij} do. We denote the contribution from e_{ij} to the already existing conflict by k_{ij}, i.e.,

$$k = \sum_{j=2}^{n} \sum_{i=1}^{j-1} k_{ij}.$$

Let us look at what happens when we bring in the specific evidence e_{ij} to the combination. As mentioned earlier this may give a conflict based on earlier evidences.

Let $S_{ij} = < x_i, x_{i+1}, \ldots, x_j >$ where $x_i = x_j = 1$ and $x_k = 0$, $i < k < j$. e_{ij} speaks against the subpath S_{ij} to the degree q_{ij}. The earlier evidences speaks in favor of this subpath to the degree

$$\frac{\text{Sup}^*(S_{ij})}{(1 - q_{ij})},$$

where $\mathrm{Sup}^*(S_{ij})$ is calculated as described earlier, but the computation of $\mathrm{Sup}^*(S_{ij})$ was then based on the evidence e_{ij} itself, which is not relevant here. The influence of e_{ij} on $\mathrm{Sup}^*(S_{ij})$ is neglected by division with its contribution $(1 - q_{ij})$. The total conflict caused by e_{ij} is consequently $c_{ij} = q_{ij} \cdot \mathrm{Sup}(S_{ij})/(1 - q_{ij})$ but this conflict is not equal to the contribution k_{ij} because a part of c_{ij} is already taken into account by the calculated conflict based on the earlier evidences. This means that c_{ij} has to be reduced in the following way. The total conflict before e_{ij} is

$$\sum_{h=2}^{j-1} \sum_{k=1}^{h-1} k_{kh} + \sum_{k=i+1}^{j-1} k_{kj}.$$

This expression can be written as a sum of the four terms:

$$\sum_{h=2}^{i-1} \sum_{k=1}^{h-1} k_{kh} + \sum_{k=1}^{i-1} k_{ki} + \sum_{h=i+1}^{j-1} \sum_{k=1}^{h-1} k_{kh} + \sum_{k=i+1}^{j-1} k_{kj}.$$

Let us consider the first term:

$$\sum_{h=2}^{i-1} \sum_{k=1}^{h-1} k_{kh}.$$

This conflict is only based on the evidences concerning vertices before v_i, therefore we may have a conflict based on these evidences at the same time as we have a conflict only based on evidences from vertex v_i and forward. The new contribution to the conflict, k_{ij}, must not contain the earlier conflict. Hence, c_{ij} is reduced by the term

$$c_{ij} \cdot \sum_{h=2}^{i-1} \sum_{k=1}^{h-1} k_{kh},$$

which is the degree to which we have conflict in both.

For the second term,

$$\sum_{k=1}^{i-1} k_{ki},$$

the reasoning is almost the same as for the first term with the difference that in the expression for the simultaneous conflict,

$$c_{ij} \cdot \sum_{k=1}^{i-1} k_{ki},$$

the support p_i for the evidence e_i occurs in both the factor c_{ij} and the factors k_{ij}, which must not be the case when they are regarded simultaneously, therefore the expression has to be divided by p_i, i.e., the reducing term based on:

$$\sum_{k=i+1}^{j-1} k_{ki},$$

equals:

$$\frac{c_{ij}}{p_i} \cdot \sum_{k=i+1}^{j-1} k_{ki}.$$

For the last two terms in the sum above, every k_{kh} is based on at least one evidence e_k concerning a vertex between v_i and v_j. This means that it is impossible to have a conflict based on the evidence e_{ij} at the same time as we state a vertex between v_i and v_j to be true, so the last two terms in the sum do not contain any part of the conflict c_{ij} and do not contribute to the reduction.

This means that

$$\begin{aligned} k_{ij} &= c_{ij} \cdot \left(1 - \sum_{h=2}^{i-1} \sum_{k=1}^{h-1} k_{kh} - \frac{1}{p_i} \cdot \sum_{k=1}^{i-1} k_{ki}\right) \\ &= \frac{q_{ij}}{(1 - q_{ij})} \cdot \left(1 - \sum_{h=2}^{i-1} \sum_{k=1}^{h-1} k_{kh} - \frac{1}{p_i} \cdot \sum_{k=1}^{i-1} k_{ki}\right). \end{aligned}$$

This is true for $i \geq 3$.

If $i = 1$ the c_{ij} do not have to be reduced because in this case the reasoning is the same as for the last two terms.

For $i = 2$ the reducing factor is:

$$\frac{1}{p_i} \cdot \sum_{k=1}^{i-1} k_{ik},$$

and

$$k_{ij} = \frac{1}{1 - q_{ij}} \cdot \left(1 - \frac{1}{p_i} \cdot \sum_{k=1}^{i-1} k_{ki}\right).$$

6.3.2. THE ALGORITHM FOR CONFLICT The conflict, k_n, of a graph with n vertices can be calculated as

$$k_n = \begin{cases} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_{ij}, & n > 1 \\ 0, & n = 1 \end{cases}$$

where

$$k_{ij} = \begin{cases} \left(k_{1j-i+1}\sigma_{i-1}^{\mathrm{I}}\right)\sigma_i^{\mathrm{II}}, & i > 1 \\ \dfrac{q_{1j}}{1-q_{1j}} \cdot \mathrm{Spt}^*(< r_1, \neg r_2, \neg r_3, \ldots, \neg r_{j-1}, r_j >), & i = 1 \end{cases}$$

and σ_i^{I} and σ_i^{II} are the substitutions

$$\sigma_i^{\mathrm{I}} = \forall m, n.\{p_m/p_{m+i}, q_{mn}/q_{m+i\,n+i}\},$$

$$\sigma_i^{\mathrm{II}} = \begin{cases} \left\{p_i \Big/ \left(p_i - p_i \cdot \sum_{m=1}^{i-2} \sum_{n=m+1}^{i-1} k_{mn} - \sum_{m=1}^{i-1} k_{mi}\right)\right\}, & i > 2 \\ \{p_i/(p_i - k_{12})\}, & i = 2 \end{cases}$$

and $\mathrm{Spt}^*(< r_1, \neg r_2, \neg r_3, \ldots, \neg r_{j-1}, r_j >)$ is the unnormalized support.

7. COMPLEXITY

The time complexity of the classic algorithm is of course such that using it in any real-time application is out of the question. But even when one is using it for symbolic precalculations one runs into problems, as seen in Figure 12. The space complexity of the classic algorithm, Figure 12, should, however, not be interpreted as the size of the data to be handled by an application, but rather the size of the expressions that ought to be simplified by some algebraic system.

Neither can the new algorithm be used in real-time applications for anything but the smallest problems, but it is feasible to use it for other applications as well as for symbolic precalculations. On today's supercomputers the new algorithm can manage graphs of up to 36 vertices in size, i.e., up to 666 evidences with $|\Theta| = 2^{36}$, when calculating support for one

	The classic algorithm	**The new algoritm**
Time complexity	$O(\lvert\Theta\rvert^{\log\lvert\Theta\rvert})$	$O(\lvert\Theta\rvert \cdot \log\lvert\Theta\rvert)$
Space complexity	$O(\lvert\Theta\rvert^{\log\lvert\Theta\rvert} \cdot \log^2\lvert\Theta\rvert)$	$O(\lvert\Theta\rvert \cdot \log\lvert\Theta\rvert)$

Figure 12. Complexity of the classic and new algorithms.

single instance of the frame (1 Gflops for 10 minutes) as compared to only six vertices for the classic algorithm.

7.1. The Classic Algorithm

Assuming that there are n vertices in the graph, the time complexity of Bel_p is $O(2^n)$ and the space complexity is $O(n \cdot 2^n)$. When there are n vertices there are $\frac{1}{2} \cdot n \cdot (n-1)$ edges. Thus, the time complexity of Bel_n becomes $O(2^{(n^2)})$ with a space complexity of $O(n^2 \cdot 2^{(n^2)})$. The time complexity of $\mathrm{Bel}_p \oplus \mathrm{Bel}_n$ will then be $O(2^{(n^2)})$ and the space complexity $O(n^2 \cdot 2^{(n^2)})$, or when measured in the size of the frame $O(|\Theta|^{\log|\Theta|})$ and $O(|\Theta|^{\log|\Theta|} \cdot \log^2|\Theta|)$ respectively.

7.2. The New Algorithm

The unnormalized plausibility for a single path can be calculated in linear time. The time complexity of the unnormalized support for a single path is far worse, being determined by the summation over the three last factors that are $O(n \cdot (\frac{3}{2})^n)$, $O(n \cdot (\frac{3}{2})^n)$ and $O(n \cdot 2^n)$ respectively. Thus, the time complexity of calculating the unnormalized support for a single instance of the frame becomes $O(n \cdot 2^n)$. If we assume that the unnormalized support for one particular path for each graph size is already calculated, then the time complexity of calculating the conflict will be $O(2^n)$, otherwise we must calculate the unnormalized support for these paths yielding a time complexity for the conflict of $O(n \cdot 2^n)$. Thus, the time complexity of calculating support and plausibility for each path is $O(n \cdot 2^n)$, or when measured in the size of the frame $O(|\Theta| \cdot \log|\Theta|)$. Presumably we can use domain knowledge to substantially restrict the number of credible scenarios.

The space complexity, when calculating support and plausibility symbolically, is equal to the time complexity.

8. CONCLUSIONS

We have presented an algorithm that makes it feasible to precalculate support and plausibility symbolically for completely specified paths through a complete directed acyclic graph. One problem when reasoning about completely specified paths, i.e., paths where $\forall i.\ x_i \neq \theta_i$, is that for larger graphs there might be a large number of quite similar paths with equally low support and plausibility. The average characterization of these paths may then be lost. If there is no completely specified path that stands out from the analysis, this would make the calculation useless for decision support. We might therefore also be interested in reasoning about incompletely specified paths, i.e., subparts of paths.

ACKNOWLEDGMENTS

We would like to thank Stefan Arnborg and Per Svensson for their helpful comments regarding this paper.

References

1. Dempster, A. P., A generalization of Bayesian inference, *J. Roy. Stat. Soc. Ser. B* 30(2), 205–247, 1968.
2. Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N.J., 1976.
3. Shafer, G., Perspectives on the theory and practice of belief functions, *Int. J. Approx. Reasoning* 4(5/6), 323–362, 1990.
4. Shafer, G., Rejoinders to comments on "Perspectives on the theory and practice of belief functions," *Int. J. Approx. Reasoning* 6(3), 445–480, 1992.
5. Lowrance, J. D., Garvey, T. D., and Strat, T. M., A framework for evidential-reasoning systems, *Proc. 5th Nat. Conf. AI*, Philadelphia, PA, 896–901, 1986.
6. Schubert, J., On nonspecific evidence, *Int. J. Intell. Syst.*, to appear.
7. Bergsten, U, Schubert, J., and Svensson, P., Dezzy-Ett Demonstrationssystem för Analys av Ubåtsunderrättelser, in Delprojekt Informationssystem inom Huvudprojekt Ubåtsskydd-Slutrapport (Schubert, J., Ed.), FOA report A 20046-2.7, Department of Weapon Systems, Effects and Protection, National Defence Research Establishment, Stockholm, 8–10, 1990.
8. Kennes, R., and Smets, P., Computational Aspects of the Möbius Transformation, in *Uncertainty in Artificial Intelligence* 6, Machine Intelligence and Pattern Recognition Ser., Vol. 12 (P. P. Bonissone, M. Henrion, L. N. Kanal and J. F. Lemmer, Eds.), Elsevier Science Publ. (North-Holland), Amsterdam, 401-416, 1991.
9. Thoma, H. M., *Factorization of Belief Functions*, Doctoral Dissertation, Dept. of Statistics, Harvard Univ., Cambridge, MA,1989.
10. Gordon, J., and Shortliffe, E. H., A method for managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intell.* 26(3), 323–357, 1985.
11. Barnett, J. A., Computational Methods for a Mathematical Theory of Evidence, *Proc. 7th Int. Joint Conf. AI*, Vancouver, BC, 868–875, 1981.
12. Shafer, G., and Logan, R., Implementing Dempster's rule for hierarchical evidence, *Artificial Intell.* 33(3), 271–198, 1987.

13. Shafer, G., Shenoy, P. P., and Mellouli, K., Propagating belief functions in qualitative Markov trees, *Int. J. Approx. Reasoning* 1(4), 349–400, 1987.

14. Mellouli, K., Shafer, G., and Shenoy, P. P., Qualitative Markov Networks, in *Uncertainty in Knowledge-Based Systems*, Lecture Notes in Computer Science, Vol. 286 (B. Bouchon and R. R. Yager, Eds.), Springer–Verlag, Berlin, 69–74, 1987.

15. Shenoy, P. P., and Shafer, G., Axioms for Probability and Belief-Function Propagation, in *Uncertainty in Artificial Intelligence* 4, Machine Intelligence and Pattern Recognition Ser., Vol. 9 (R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, Eds.), Elsevier Science Publ. (North-Holland), Amsterdam, 169–198, 1990.

On Rho in a Decision-Theoretic Apparatus of Dempster-Shafer Theory

Manuscript

V

On Rho in a Decision-Theoretic Apparatus of Dempster-Shafer Theory

Johan Schubert
Division of Information System Technology, Department of Command and Control Warfare Technology, National Defence Research Establishment, Stockholm, Sweden.

ABSTRACT

Thomas M. Strat has developed a decision-theoretic apparatus for Dempster-Shafer theory (Decision analysis using belief functions, Int. J. Approx. Reasoning 4(5/6), 391-417, 1990). In this apparatus, expected utility intervals are constructed for different choices. The choice with the highest expected utility is preferable to others. However, to find the preferred choice when the expected utility interval of one choice is included in that of another, it is necessary to interpolate a discerning point in the intervals. This is done by the parameter ρ, *where* ρ *is defined as the probability that the ambiguity about the utility of every non-singleton focal element will turn out as favorable as possible. If there are several different decision makers we might sometimes be more interested in having the highest expected utility among the decision makers rather than only trying to maximize our own expected utility regardless of choices made by other decision makers. The preference of each choice is then determined by the probability of yielding the highest expected utility. This probability is equal to the maximal interval length of* ρ *under which an alternative is preferred. We must here take into account not only the choices already done be other decision makers but also the rational choices we can assume to be made by later decision makers. In Strats apparatus, an assumption, unwarranted by the evidence at hand, has to be made about the value of* ρ. *In this article we demonstrate that no such assumption is necessary. It is sufficient to assume a uniform probability distribution for* ρ *to be able to discern the most preferable choice. We will discuss when this approach is justifiable.*

KEYWORDS: ***belief functions, decision analysis, decision making, decision tree, Dempster-Shafer theory, evidential reasoning, reasoning under uncertainty***

Address correspondence to Johan Schubert, Division of Information System Technology, Department of Command and Control Warfare Technology, National Defence Research Establishment, S-172 90 Stockholm, Sweden.

1. INTRODUCTION

To make rational decisions under uncertainty is somewhat complicated in Dempster-Shafer theory (Dempster [1], Shafer [2]) because of the interval representation. In [3] Nguyen and Walker discussed different approaches to decision making with belief functions. They found three different basic models. The first is based on the Choquet integral that yields the expected utility with respect to belief functions;

$$E_F(u) = \int_0^{\infty} F(u > t)dt + \int_{-\infty}^{0} [F(u > t) - 1]\, dt$$

where F is a belief function defined on 2^{Θ} by $F(A) = inf\{P(A): P \in \mathbf{P}\}$ and $\mathbf{P} = \{P: F \leq P\}$ is a class of probability measures on Θ. This leads to the pessimistic strategy of ranking alternatives by their minimal expected utility.

In the second basic model the decision maker uses some additional information or subjective views. Instead of searching for the alternative that maximizes expected utility the utility function will be supplemented by some new function dependent on the utility and some other parameter corresponding to the additional information or subjective views. An article by Strat [4] is an example of the second basic model.

The third basic model consists of models using the insufficient reason principle or equivalently the maximum entropy principle.

As an example, Smets and Kennes [5] have developed a two-level model of credal belief and pignistic probability, called the 'Transferable belief model' (TBM), that belongs to the third category of Nguyen and Walker [3].

On the credal level of this model the reasoning process takes place in the usual manner as within Dempster-Shafer theory. Here beliefs are held by belief functions and combined by Dempster's rule.

When a decision must be taken, the belief on the credal level is transformed to a probability at the pignistic level by a "pignistic transformation" based on Laplace's insufficient reasoning principle;

$$BetP(x) = \sum_{x \subseteq A \in \mathfrak{R}} \frac{m(A)}{|A|} = \sum_{A \in \mathfrak{R}} m(A) \cdot \frac{|x \cap A|}{|A|},$$

$$BetP(B) = \sum_{A \in \mathfrak{R}} m(A) \cdot \frac{|B \cap A|}{|A|}$$

where $BetP(\cdot)$ is the pignistic probability we should use to 'bet' with in a utility maximization process. Here $\mathfrak{R}$ is the set of all propositions. It is called the betting frame.

The pignistic probability regarding some proposition A depends on the organization of the betting frame $\Re$. But regardless of the organization of the betting frame we always have $BetP(A) \geq Bel(A) \quad \forall A \in \Re$.

Further discussions on decision making with belief functions can be found in [6, 7].

This article is concerned with a method that has recently been developed by Strat [4]. In this method an expected utility interval is constructed for each choice;

$$[E_*(x), E^*(x)]$$

where $E_*(\cdot)$ and $E^*(\cdot)$ are defined as

$$E_*(x) \triangleq \sum_{A_i \subseteq \Theta} inf(A_i) \cdot m_\Theta(A_i)$$

and

$$E^*(x) \triangleq \sum_{A_i \subseteq \Theta} sup(A_i) \cdot m_\Theta(A_i),$$

Θ is a frame of discernment, i.e., an exhaustive set of mutually exclusive possibilities, and m_Θ is a basic probability assignment, a function from the power set of Θ to [0, 1]:

$$m_\Theta: 2^\Theta \rightarrow [0,1]$$

whenever

$$m_\Theta(\varnothing) = 0$$

and

$$\sum_{A_i \subseteq \Theta} m_\Theta(A_i) = 1.$$

We will call E_* the lower expected utility and E^* the upper expected utility.

Our preference among different alternatives will depend upon their expected utility. Let the expected utility be defined as

$$E(x) \overset{\Delta}{=} E_*(x) + \rho \cdot (E^*(x) - E_*(x))$$

where ρ is defined as the probability that the ambiguity about the utility of every non-singleton focal element will turn out as favorably as possible, i.e. the probability that nature will turn out as favorably as possibly towards us as decision makers. This article will establish an alternative to making an outright, and often unwarranted, assumption about ρ. This alternative is to accept a uniform probability distribution for ρ.

Adopting a uniform probability distribution for ρ requires two conditions being fulfilled. Firstly, there certainly must not be any evidence at hand regarding the value of the probability ρ. Such evidence could, for example, be in the form of domain knowledge, direct evidence regarding the value of ρ or knowledge that the decision situation is controlled by either the decision maker or an adversary. It would seem to be commonplace that there is no direct evidence available regarding the value of ρ. The situation we are looking for is then a business like situation in a field with poor domain knowledge where the outcomes are not controlled by either the decision maker or an adversary, i.e. a decision situation without evidence regarding the value of ρ. Secondly, it must be a decision situation where the decision maker is not only interested in minimizing the expected loss regardless of the possible gains or interested in maximizing the expected gain regardless of the possible losses. In these two situations he would choose to adopt $\rho = 0$ and $\rho = 1$, respectively, even if there is no available evidence regarding the value of ρ. This would be the situation if the decision maker is forced to play a game he thinks is unfavorable. Then he would try to minimize the expected loss, i.e. choosing $\rho = 0$. If, on the other hand, the decision maker is forced to obtain a lot of value by playing a particular game, he may try to maximize the expected gain, i.e. choosing $\rho = 1$. This eliminates the extreme situations where the decision maker is forced into a game by one reason or another, i.e. situations where it is not possible to make no choice. What is remaining are the "normal" business like decision situations where we do not have a reason to choose one value for ρ over another when there is not any evidence at hand regarding the value of ρ.

As Strat points out in his article, if we make an assumption about the value of ρ we should not confuse our assumption about ambiguity with our risk preference. Our risk preference is handled by adopting utilities.

The methodology in this article was developed as the decision part of a multiple-target tracking algorithm (Schubert [8], Bergsten and Schubert [9]) for an anti-submarine intelligence analysis system.

In Section 2 we will discuss points of preference change and in Section 3 the uniform probability distribution for ρ. In Section 4 we will study decision making with a uniform probability distribution for ρ and the different objectives decision makers might have. Finally, conclusions are drawn in Section 5.

2. THE PREFERRED CHOICE

Obviously, when we are searching for the most preferable choice we can immediately disregard those choices where the upper expected utility is less than the highest lower expected utility among all choices. Furthermore, if both interval limits of the utility interval are higher for one alternative than for another, i.e. $E_{i*} > E_{j*}$ and $E_i^* > E_j^*$, then this one, choice *i*, is always preferable regardless of the value of ρ. In fact, if we receive the choices ordered by falling magnitude of their upper expected utility we can immediately disregard any choice whose lower expected utility is less than any lower expected utility of the previous choices. Only if the expected utility interval of one choice is included in the interval of another choice will our preference depend on the assumed value of ρ. As a result, we will end up with a set of expected utility intervals ordered by interval inclusion, $[E_{1*},E_1^*] \subseteq [E_{2*},E_2^*] \subseteq \ldots \subseteq [E_{n*},E_n^*] \subseteq [0,1]$. Here we have renumbered the choices by the order of interval inclusion, i.e. by order of increasing interval length. In the following we will only consider choices ordered and renumbered by interval inclusion.

Let us study the choice between x_1 and x_2 where $[E_{1*},E_1^*] \subseteq [E_{2*},E_2^*]$;

$$\textit{Choice 1: } [E_{1*},E_1^*],$$

$$\textit{Choice 2: } [E_{2*},E_2^*].$$

Here choice 1 is preferred when

$$E_{1*} + \rho \cdot (E_1^* - E_{1*}) > E_{2*} + \rho \cdot (E_2^* - E_{2*})$$

We find that both choices are equally preferable if

$$\rho = \frac{E_{1*} - E_{2*}}{(E_2^* - E_{2*}) - (E_1^* - E_{1*})}.$$

Let us call this value ρ_{12}. Since choice 1 has the highest lower expected utility of the two choices it is preferred when $\rho \in [0,\rho_{12}]$, and choice 2 is preferred when $\rho \in [\rho_{12},1]$.

When we have several choices they may be preferred in different intervals of ρ. If we calculate all ρ_{ij}'s and order them by increasing magnitude we can calculate the expected utility of every choice for a point in each interval of the ordered ρ_{ij}'s. The choice with the highest expected utility in each interval is then the preferred choice for that interval. However, we already know that choice 1 is preferred when $\rho = 0$, since this choice has the highest lower expected utility among all choices, and it will remain the preferred choice while ρ is less than $\min_i \rho_{1i}$, the smallest of all ρ_{ij}'s and the first point of preference change. Beyond this point, choice i will be preferable over choice 1. Since choice 1 will never again be preferred in any other interval, we may now disregard all other ρ_{1j}, $j \neq i$, even though they represent points of possible preference change. The reason for this is obvious, choice 1 can never again be the most preferable choice for any interval above $\min_i \rho_{1i}$ since it is not even preferred to choice i beyond that point. Thus, these points of possible preference change will never represent an actual change of the current preference. Continuing, choice i will now be preferred up to the point where $\rho = \min_j \rho_{ij}$, and beyond this point choice j will be preferred up until $\rho = \min_k \rho_{jk}$, etc. Thus, by iteration we find that the choices are each preferable in the following intervals:

$$\textit{Choice } 1\textit{:}\ [0, \min_i \rho_{1i}],$$

$$\textit{Choice } i\textit{:}\ [\min_i \rho_{1i}, \min_j \rho_{ij}],$$

$$\textit{Choice } j\textit{:}\ [\min_j \rho_{ij}, \min_k \rho_{jk}],$$

$$\ldots$$

$$\textit{Choice } n\textit{:}\ [\rho_{mn}, 1].$$

Alternatively, for any choice j that is preferable somewhere, its interval of preference can be described as

$$\textit{Choice } j\textit{:}\ [\max_i \rho_{ij}, \min_k \rho_{jk}].$$

If two or more ρ_{ij}'s are equal in a minimization, $\min_j \rho_{ij}$, the next preferred choice would be ambiguous. In this case we take the choice with the highest number. If not, we would end up with one or more choices preferred under a zero interval length of ρ before we would get this choice anyway.

3. A UNIFORM PROBABILITY DISTRIBUTION FOR ρ

All we know about the value of ρ is that it is a parameter the belongs to the set of real numbers between 0 and 1, $\rho \in [0, 1]$, i.e. we know that our frame is that same set of numbers, $\Theta = [0, 1]$. Thus, apart from knowing the frame for ρ we do not know anything at all. We have a vacuous bpa where $m(\Theta) = 1$. In order not to reduce the overall nonspecificity of this initial state when making an assumption about the probability distribution about ρ we might ask that any such assumption about ρ should yield the same nonspecificity as what we have now. We define the nonspecificity as

$$\mathrm{I}(m) = \sum_{A \in F} m(A) \cdot Log_2|A|$$

which is a generalization of Hartley's information [10].

Calculating the nonspecificity $\mathrm{I}(m)$ of this initial state where $F = \{\Theta\}$ and $m(\Theta) = 1$ we have

$$\begin{aligned}\mathrm{I}(m) &= \sum_{A \in F} m(A) \cdot Log_2|A| \\ &= m(\Theta) \cdot Log_2|\Theta| = 1 \cdot Log_2|\Theta|,\end{aligned}$$

and since Θ is the infinite set of real numbers between 0 and 1 we receive an infinite nonspecificity.

If we make a single-point assumption about ρ where $F = \{\{\rho\}\}$ and $m(\{\rho\}) = 1$ we receive a nonspecificity of

$$\begin{aligned}\mathrm{I}(m) &= \sum_{A \in F} m(A) \cdot Log_2|A| \\ &= m(\{\rho\}) \cdot Log_2|\{\rho\}| = 1 \cdot Log_2 1 = 0,\end{aligned}$$

and for any point-wise distribution for ρ where $F = \{\{\rho_1\}, \{\rho_2\}, \ldots\}$ we get

$$\begin{aligned}\mathrm{H}(m) &= \sum_{A \in F} m(A) \cdot Log_2|A| \\ &= \sum_{A \in F} m(A) \cdot Log_2 1 = 0.\end{aligned}$$

Obviously, our distribution needs a continuous part to reach the infinite nonspecificity of the initial state. Any such distribution with just one continuous part, B, will reach infinite nonspecificity. We have

$$\begin{aligned}\mathrm{I}(m) &= \sum_{A \in F} m(A) \cdot Log_2|A| \\ &= m(B) \cdot Log_2|B|\end{aligned}$$

where $F = \{B, \{\rho_1\}, \{\rho_2\}, \ldots\}$ and B is an interval of real numbers included in [0, 1]. If B is of infinite size we have an infinite nonspecificity.

Furthermore, we might also demand that the nonspecificity of our new distribution should be equal to the original assignment for any size of the frame. Let $F = \{B_1, B_2, \ldots, \{\rho_1\}, \{\rho_2\}, \ldots\}$ where B_i's are intervals included in [0, 1]. We must then have

$$Log_2|\Theta| = \sum_{A \in F} m(A) \cdot Log_2|A|.$$

Here $A \subseteq \Theta$ and thus we may write $|A| = \alpha_A \cdot |\Theta|$ where

$$\frac{1}{|\Theta|} \leq \alpha_A \leq 1$$

and

$$\alpha_{\{\rho_i\}} = \frac{1}{|\Theta|}.$$

We have

$$Log_2|\Theta| = \sum_{A \in F} m(A) \cdot Log_2(\alpha_A \cdot |\Theta|) = \sum_{A \in F} m(A) \cdot Log_2\alpha_A + \sum_{A \in F} m(A) \cdot Log_2|\Theta|$$
$$= \sum_{A \in F} m(A) \cdot Log_2\alpha_A + Log_2|\Theta| \cdot \sum_{A \in F} m(A).$$

From this it immediately follows that

$$\sum_{A \in F} m(A) \cdot Log_2\alpha_A = 0.$$

Since $m(A) > 0$ for every A and $Log_2\alpha_A \leq 0$ for every α_A, we must have that $\alpha_A = 1$ for every A. But since $|A| = \alpha_A \cdot |\Theta|$, and Θ is the entire frame it follows that $A = \Theta$, i.e. that we have only one focal element $F = \{\Theta\}$.

This means that we have only one continuous part of the probability distribution for ρ, and that it covers the entire interval from 0 to 1, i.e. a uniform probability distribution.

4. DECISION MAKING

4.1. Decision making with a uniform probability distribution for ρ

If we refrain from making an unwarranted assumption about the value of ρ we might instead accept a uniform probability distribution for ρ, i.e. the assumption that all values of ρ are equally probable. Any of the above choices that are preferable somewhere might now be preferred. However, the probability for the choices to be preferred are not equal. This probability varies with the length of the interval under which it is preferred.

If we are only interested in simple maximizing of utility then adopting a uniform probability distribution for ρ yields the same result as setting $\rho = 0.5$. Then, for simplicity, we might as well set $\rho = 0.5$ and choose the alternative that yields the highest expected utility as our decision.

However, in a situation with several different decision makers we might sometimes be more interested in having the highest expected utility among the decision makers rather than only trying to maximize our own expected utility.

Thus, rather than actually making a random assumption about ρ in order to find a preferable choice it makes sense to prefer the choice that is most likely preferred *if* the value of ρ was determined at random. Assuming the uniform probability distribution for ρ, this is obviously the choice that is preferred under the maximum interval length of ρ. This might be according to the principle "it is better to choose what is most likely the best alternative rather than to gamble for it." The interval length under which a choice is preferred, Pref(·), is here defined as

$$\mathrm{Pref}(x_j) \triangleq max(0, min_k\ \rho_{jk} - max_i\ \rho_{ij})$$

where $\min_k \rho_{nk} \triangleq 1$ and $\max_i \rho_{i1} \triangleq 0$.

If the number of alternatives is equal to the number of decision makers then all we have to do is to choose the alternative that is preferred under the maximal interval length. That will be the choice with the highest probability of giving us the highest expected utility.

The situation becomes more complex when the number of decision makers are less than the number of choices.

4.2. An example

Let us consider an example with four choices whose expected utility intervals are ordered by interval inclusion:

Choice 1*:* [0.5, 0.6],

Choice 2*:* [0.4, 0.7],

Choice 3*:* [0.3, 0.9],

Choice 4*:* [0.2, 1.0].

Calculating the points of preference change gives us

$$\rho_{12} = \frac{E_{1*} - E_{2*}}{(E_2^* - E_{2*}) - (E_1^* - E_{1*})} = \frac{0.5 - 0.4}{(0.7 - 0.4) - (0.6 - 0.5)} = 0.5$$

and by the same formula $\rho_{13} = 0.4$, $\rho_{14} = 0.43$, $\rho_{23} = 0.33$, $\rho_{24} = 0.4$, $\rho_{34} = 0.5$. We find by iteration that the choices are preferable in the following intervals of ρ:

$$\textit{Choice } 1\text{: } [0, min_i\ \rho_{1i}] = [0, \rho_{13}] = [0, 0.4],$$

$$\textit{Choice } 3\text{: } [0.4, min_j\ \rho_{3j}] = [0.4, \rho_{34}] = [0.4, 0.5],$$

$$\textit{Choice } 4\text{: } [0.5, 1],$$

and being preferred under the following interval lengths:

$$Pref(x_1) = 0.4,$$

$$Pref(x_2) = 0,$$

$$Pref(x_3) = 0.1,$$

$$Pref(x_4) = 0.5.$$

In this case choice 2 will never be preferred regardless of the value of ρ. If an unwarranted assumption is made about ρ any of the other three choices could be preferred. If we, on the other hand, only assume a uniform probability distribution for ρ, choice 4 will be considered preferable since it is preferred under the maximum interval length of ρ.

4.3. An algorithm for finding the preferred choice

We may now find the preferred choice given a uniform probability distribution by the following algorithm.

Algorithm. *Let S be the empty set.*

1. *Order and renumber all choices by falling magnitude of upper expected utility.*

2. *For* $i = 1$ *to* n *do*

   2.1. *Add all choices whose expected utility interval belongs to set of intervals ordered by interval inclusion; If* $E_{i*} > E_{i-1*}$ *then* $S := S + \{[E_{i*}, E^*_i]\}$.

3. *Renumber all choices in S by order of increasing interval length magnitude.*

4. *For all combinations of pairs in S calculate*

$$\rho_{ij} = \frac{E_{i*} - E_{j*}}{(E_j^* - E_{j*}) - (E_i^* - E_{i*})}.$$

5. ρ_c := 0, *i* := 1, *maximum_preference* := 0.

6. *Calculate the intervals of preference for each choice and find the most preferable choice;*
While $i \neq n$ *do*

6.1. $\rho'_c := min_j\ \rho_{ij}$, *where* $min_k\ \rho_{nk} \triangleq 1$.

6.2. *Pref*(x_i) = ρ'_c - ρ_c.

6.3. *If Pref*(x_i) > *maximum_preference then*

6.3.1. *maximum_preference* := *Pref*(x_i), *preferred_choice* := *i*.

6.4. *i* := *j*.

6.5. ρ_c - ρ'_c

7. *Answer preferred_choice.*

4.4. Possible refinements

Instead of changing from the strongest possible assumption of a point-value for ρ to the weakest possible assumption of a uniform probability distribution, we may occasionally have a reason to assume some other probability distribution for ρ (Figure 1).

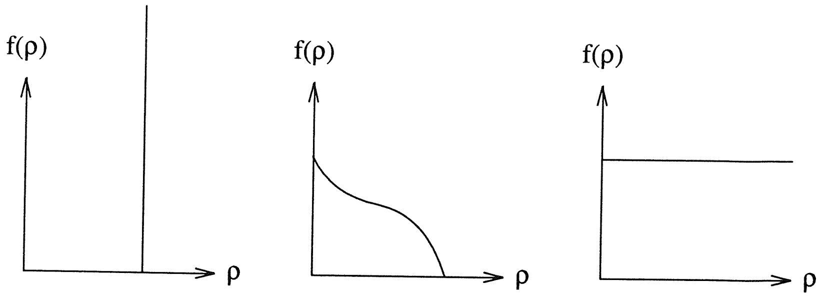

Fig. 1. A point-valued, arbitrary and uniform probability distribution for ρ.

We might for instance have some knowledge regarding a lower and upper bound for ρ. Let us call these bounds lower ambiguity probability, ρ_*, and upper ambiguity probability, ρ^*, respectively. These bounds force a simple change in the definition of preference, Pref(·);

$$\mathrm{Pref}(x_j) \triangleq max(0, min(\rho^*, min_k \ \rho_{jk}) - max(\rho_*, max_i \ \rho_{ij}))$$

where $\min_k \rho_{nk} \triangleq 1$ and $\max_i \rho_{i1} \triangleq 1$.

To incorporate the new definition of preference into the algorithm we make the following change in step 6.2.,

$$6.2. \ Pref(x_i) = max(0, min(\rho^*, \rho'_c) - max(\rho_*, \rho_c)),$$

giving all choices preferred in intervals outside the bounds of lower and upper ambiguity probability a preference of zero.

Obviously, we must be able to assume any probability distribution for ρ, $f(\rho)$. We can make a general definition of preference as

$$Pref(x_j) \triangleq max(0, \int_{max_i \ \rho_{ij}}^{min_k \ \rho_{jk}} f(\rho) d\rho)$$

where $\min_k \rho_{nk} \triangleq 1$ and $\max_i \rho_{i1} \triangleq 1$.

Finally, we change the computation of preference in step 6.2. of the algorithm to

$$6.2. \ \ Pref(x_j) \triangleq F(\rho'_c) - F(\rho_c)$$

where

$$F(\rho) = \int f(\rho) d\rho.$$

4.5. Two decision makers searching for the most preferable choice

When two decision makers compete for the highest utility the preference of each alternative is determined by the chance of having the alternative that is preferred under the maximal interval length of ρ after our opponent has also made his choice. If we assume we have the first choice then our opponent will make his choice taking into account the choice we made. Since our goal is to have the highest possible probability of having the best alternative we must also take into account the best choice our opponent can make. It is found by choosing the alternative with the highest preference when preference is defined as

$$\text{Pref}(x_j) \triangleq min(min_k \ \rho_{jk}, 1 - max_i \ \rho_{ij}).$$

Here $min_k \ \rho_{jk}$ is the preference for choice j when our opponent chooses his best alternative k where $k > j$ and $1 - max_i \ \rho_{ij}$ is the preference for choice j when our opponent chooses his best alternative i where $i < j$.

If we, on the other hand, is the second of the two decision makers the situation is even simpler. We just have to find the choice with maximal preference where preference is defined as

$$\text{Pref}(x_j, k) = \begin{cases} \rho_{jk}, j < k \\ \rho_{kj}, j > k \end{cases}$$

and k is the alternative already chosen by our opponent.

4.6. Several decision makers

When the number of decision makers are less than the number of choices the situation becomes much more complex. We must here take into account not only the choices already done be other decision makers but also the rational choices we can assume to be made by later decision makers. This is the case since the length of the preference interval for any alternative depends on the other choices that are made. If I^* is the set of all choice done by previous decision makers the preference of a choice x_j may be calculated as

$$\text{Pref}(x_j, I^*) = max(0, \underset{k \in I^* + I_*(I^*, j)}{min_k \ \rho_{jk}} - \underset{i \in I^* + I_*(I^*, j)}{max_i \ \rho_{ij}})$$

where $I_*(I^*, j)$ is the set of rational choices later decision maker will make given our choice j. For any decision maker we want to find the alternative that maximizes his preference, i.e.

$$\underset{j \in I - I^*}{max_j} \ \text{Pref}(x_j, I^*)$$

where I is the set of all possible choices.

This problem is solved starting with the final choice done by the last of the n decision makers, and for all possible sets of earlier choices I^*. Here $|I^*| = n - 1$ and $I_* = \varnothing$. We find the earlier choices by stepping backwards through all possible sequences of choices done by different decision makers until we reach the first choice done by the first decision maker.

This can be seen as going up a tree with one decision maker at each level until we reach the first decision maker at the root of the tree. Each branch at a certain level of the tree corresponds to a different sequence of choices made by the earlier decision makers. The edges going down from each node at this level corresponds to the possible choices that can be made by the decision maker at this level.

5. CONCLUSION

We have demonstrated that it is not necessary to make a point-value assumption about ρ in Strat's decision-theoretic apparatus of Dempster-Shafer theory. In fact, it is sufficient to assume a uniform probability distribution for ρ to be able to discern the most preferable choice. We give an algorithm for finding the most preferable choice based on an iterative search of points of preference change among choices ordered by interval inclusion. We discuss the ability to assume any probability distribution for ρ.

We also discuss the more complex problem of several decision makers competing for the highest expected utility. The preference of each alternative to some decision makers is shown to be the probability that the alternative has the highest expected utility after all decision makers have made their choices, where we take into account both the choices already done be other decision makers and the rational choices we can assume to be made by later decision makers.

ACKNOWLEDGMENTS

I would like to thank Stefan Arnborg, Ulla Bergsten and Per Svensson for their helpful comments regarding this article.

References

1. Dempster, A. P., A generalization of Bayesian inference, *J. Roy. Stat. Soc. Ser. B* **30**(2), 205-247, 1968.
2. Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N.J., 1976.
3. Nguyen, H. T., and Walker, E. A., On decision making using belief functions, in *Advances in the Dempster-Shafer Theory of Evidence* (R. Y. Yager, M. Fedrizzi and J. Kacprzyk, Eds.), John Wiley & Sons, Inc., New York, N.Y., 311-330, 1994.
4. Strat, T. M., Decision analysis using belief functions, *Int. J. Approx. Reasoning* **4**(5/6), 391-417, 1990.
5. Smets, P., and Kennes, R., The transferable belief model, *Artif. Intell.* **66**(2), 191-234, 1994.

6. Shafer, G., and Pearl, J., *Readings in Uncertain Reasoning,* The Morgan Kaufmann Ser. in Representation and Reasoning, Morgan Kaufmann Publ., San Mateo, Calif., 1990.
7. Yager, R. Y., Fedrizzi, M., and Kacprzyk, J. *Advances in the Dempster-Shafer Theory of Evidence,* John Wiley & Sons, Inc., New York, N.Y., 1994.
8. Schubert, J., On nonspecific evidence, *Int. J. Intell. Syst.* **8**(6), 711-725, 1993.
9. Bergsten, U., and Schubert, J., Dempster's rule for evidence ordered in a complete directed acyclic graph, *Int. J. Approx. Reasoning* **9**(1), 37-73, 1993.
10. N.R. Pal, J.C. Bezdek, and R. Hemasinha, Uncertainty Measures for Evidential Reasoning I: A Review, *Int. J. Approx. Reasoning* **7**(3/4), 165-183, 1992.

ISBN 91-7170-801-4

REPRO PRINT AB
Stockholm 1994