Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information Fusion 5 (2004) 309–318
INFORMATION FUSION
www.elsevier.com/locate/inffus

Clustering belief functions based on attracting and conflicting metalevel evidence using Potts spin mean field theory

Johan Schubert *

Department of Data and Information Fusion, Division of Command and Control Systems, Swedish Defence Research Agency, SE-172 90 Stockholm, Sweden

Received 5 May 2003; received in revised form 4 December 2003; accepted 4 December 2003
Available online 31 January 2004

Abstract

In this paper we develop a Potts spin neural clustering method for clustering belief functions based on attracting and conflicting metalevel evidence. Such clustering is useful when the belief functions concern multiple events, and all belief functions are mixed up. The clustering process is used as the means for separating the belief functions into clusters that should be handled independently. A measure for the adequacy of a partitioning of all belief functions is derived and mapped onto the neural network in order to obtain fast clustering. A comparison of classification error rate between using conflicting metalevel evidence only and both conflicting and attracting metalevel evidence demonstrates a significant reduction in classification error rate when using both.

Keywords: Belief function; Clustering; Potts spin; Conflict; Attraction

1. Introduction

In this paper we develop a method for handling belief functions that concern multiple events. This is the case when it is not known a priori to which event each belief function is related. The belief functions are clustered into clusters that should be handled independently.

In [2] a method for clustering belief functions based on their pairwise conflict was developed. This method was extended into a method capable of also handling pairwise attractions [15]. Such evidence is not generated intrinsically in the same way as conflicting evidence. Instead, we assume that it is given from some external source as additional information about the partitioning of the set of all belief functions. The extended method handles both types of evidence internally within all clusters.

First, in this paper, the extended method is further refined by also using conflicting and attracting metalevel evidence externally between clusters when finding the best partition of all belief functions. Secondly, the new problem formulation is mapped onto a Potts spin [20] neural network for fast clustering.

As an example let us look at a real-world problem from intelligence analysis that we are studying [4]. In intelligence analysis we may have conflicts (metalevel evidence) between two different intelligence reports about sighted objects, indicating that two objects probably do not belong to the same unit (cluster). Such conflicts arise when reports about objects are compared under the hypothesis that they refer to the same unit, e.g., report object types, times, positions and direction may be incompatible given constraints about unit structure. At the same time we may have information from communication intelligence as an external source (providing attracting metalevel evidence), indicating that the two objects probably do belong to the same unit (cluster) as they are in communication. Such information is made available from studying communication patterns, e.g., if two objects are transmitting in sequence we may calculate a probability that they are in communication and thus belong to the same unit structure.

* Tel.: +46-8-55503702; fax: +46-8-55503700.
E-mail address: schubert@foi.se (J. Schubert).

1566-2535/$ - see front matter
doi:10.1016/j.inffus.2003.12.002

As conflicts push reports apart (into different clusters) and attractions pull them together (into the same cluster), using both leads to an improved clustering result and faster computation.

Similar applications in information fusion of correlation problems using the conflict between belief function are found in [1,18,19] regarding its use in the transferable belief model [17] and applied to multi-sensor allocation in a submarine intelligence problem [1], in [8] applied to data from radar warners in a system within the Swedish Air Force, in [9] regarding mine detection when observing several mines under a multi-sensor platform, in [13] as part of a multiple processes intelligence management process, and in [7] using possibility theory.

In the terminology of Blackman [3] this methodology is *All-Neighbors Data Association* and can be labeled as *Central-Level Tracking* when applied to a tracking problem. In this paper the method is implemented as, what Blackman calls, *Deferred Decision Logic* but this can easily be changed to *Sequential Decision Logic* when applied to a sequential problem and when computation time is an issue. This was for example done in [14], extending the methodology in [2].

An alternative to the unsupervised learning of cluster memberships from conflicts originally put forward in [11] is when you assume some additionally given information such as class membership of belief functions. A recent paper [6] demonstrates how to use partial knowledge of class memberships in clustering belief functions. This method does however not consider the compatibility of the belief functions clustered into the same cluster as when using the conflict of Dempster's rule as the criterion [11]. Presumably, it would be possible to combine both ideas when such partial class memberships are available.

In Section 2 we give an introduction to the conflict in Dempster–Shafer theory, and in Section 3 we describe the problem at hand. We give an introduction to Potts spin neural networks (Section 4) and show how to map our Dempster–Shafer clustering problem onto the neural network (Section 5). Finally, in Section 6, we perform a series of simulations to compare classification errors and computation time of two different problems; one using conflicting metalevel evidence only and another using both conflicting and attracting metalevel evidence.

2. Conflict in Dempster–Shafer theory

In Dempster–Shafer theory [5,16] a problem is represented by an exhaustive set of mutually exclusive possibilities called a frame of discernment, Θ. Belief is assigned to any subset A of Θ by a mass function. The mass function is then a function from the power set of Θ to [0,1], $m : 2^{\Theta} \rightarrow [0,1]$.

If we receive two pieces of information concerning the same issue but from different sources, the two pieces can be combined to yield a more informed view. Combining two mass functions is done by calculating their orthogonal combination using Dempster's rule. Let A_i be a focal element of Bel_1 and B_j a focal element of Bel_2, i.e., A_i and B_j are subsets of Θ with $m_1(A_i) > 0$ and $m_2(B_j) > 0$, where Bel_1 and Bel_2 are belief functions. Furthermore, let e_1 and e_2 be two bodies of evidence each containing a set of propositions $\{A_i\}$ and $\{B_j\}$ and corresponding mass functions $m_1(A_i)$ for all i and $m_2(B_j)$, for all j. When we combine two belief functions we might notice that the two functions are not entirely consistent. Combining m_1 and m_2 of the two pieces of evidence may result in a conflict defined as

$$\mathrm{Conf}(\{e_1, e_2\}) = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j), \qquad (1)$$

whenever there are at least one focal element from $\{A_i\}$ and one focal element from $\{B_j\}$ such that $A_i \cap B_j = \emptyset$. For simplicity we will denote $\mathrm{Conf}(\{e_i, e_j\})$ by c_{ij} whenever the cardinality of $\{e_i\}$ is equal to two.

Thus, between a pair of belief functions the conflict is the sum of all products of support for logically inconsistent statements, e.g., if our first belief function supports A, B or C, while our second belief function concerning the same issue supports D or E, where A and D, as well as B and E, are deemed to be inconsistent statements, i.e., $A \cap D = B \cap E = \emptyset$, then the conflict between these two belief functions is $m(A)m(D) + m(B)m(E)$. This number is between zero and one. The higher this value is, the more conflict there is between the two belief functions. When the conflict is one the two belief functions are completely inconsistent, while a conflict of zero is no indication that the two belief functions belong together in some way, it is merely a lack of inconsistent information. Thus, the conflict is always a form of negative information and the lack of a negative indication is no positive information, it is just not negative. This is why we have introduced [15] pairwise attractions as a separate entity to handle the corresponding positive information.

3. Problem description

When we are reasoning under uncertainty in an environment of several different events we may find some pieces of evidence 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 this situation we must make sure that we do not by mistake combine pieces of evidence that are referring to different events.

When we have several belief functions regarding different events that should be handled independently we want to arrange them according to which event they are referring to. We partition the set of belief functions χ into clusters where each cluster χ_a refers to a particular event where belief functions within the cluster are all assumed to refer to the same event. In Fig. 1, thirteen pieces of evidence e_i are partitioned into four clusters.

However, if the belief functions are not labeled as to which event they are referring to, it is uncertain whether two different belief functions are referring to the same event and not possible to directly differentiate between two different events using only the propositions of the two belief functions.

We can then use the conflict of Dempster's rule when the two belief functions are combined, as an indication of whether they belong together. This conflict is the basis for separating belief functions into clusters. A high conflict between the two belief functions is an indication of repellency that they do not belong to the same cluster. The higher the conflict is, the less credible that they belong to the same cluster.

For each cluster we may create a new belief function on the metalevel with a proposition stating that we do not have an "adequate partition". The new belief functions do not reason about any of the original problems corresponding to the clusters. Rather they reason about the partitioning of the other belief functions into the different clusters. Just so we do not confuse the two types of belief functions, we call the new ones "metalevel evidence".

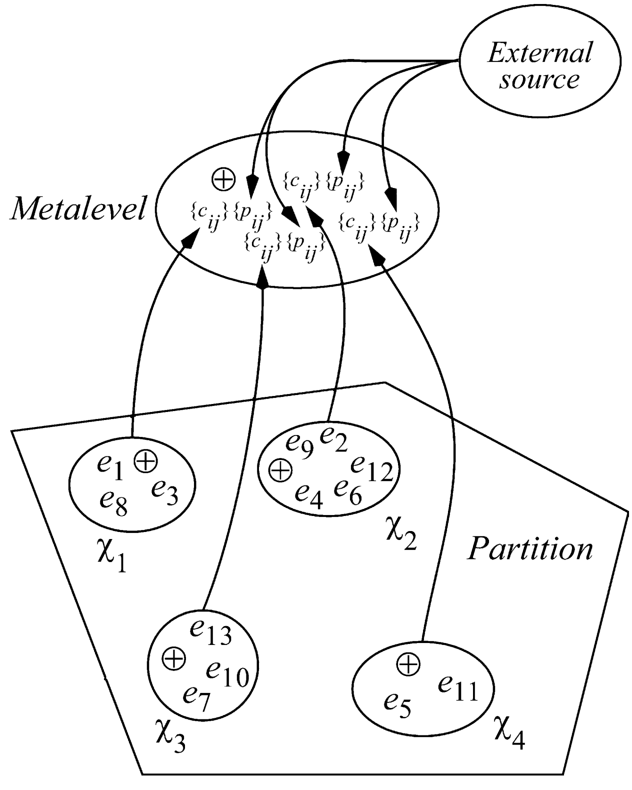

Fig. 1. The conflict in each cluster is interpreted as evidence at the metalevel, in addition we have attracting evidence from an external source.

On the metalevel we have a simple frame of discernment where $\Theta = \{\text{AdP}, \neg\text{AdP}\}$, where AdP is short for "adequate partition". Let the proposition take a value equal to the conflict of the combination within the cluster,

$$m_{\chi_a}(\neg\text{AdP}) \triangleq \text{Conf}(\{e_j | e_j \in \chi_a\}), \tag{2}$$

where $\text{Conf}(\{e_j | e_j \in \chi_a\})$ is the conflict of Dempster's rule when combining all mass functions in χ_a.

In [11] we established a criterion function of overall conflict for the entire partition called the metaconflict function (Mcf). The metaconflict was derived as the plausibility of having an adequate partitioning based on $\{m_{\chi_a}(\neg\text{AdP})\}$ for all subsets χ_a.

Definition. *Let the metaconflict function,*

$$\text{Mcf}(K, e_1, e_2, \ldots, e_N) \triangleq 1 - (1 - c_0) \cdot \prod_{a=1}^{K} (1 - c_a) \tag{3}$$

be the conflict against a partitioning of N belief functions of the set χ into K disjoint subsets χ_a, where c_a is the conflict in Dempster's rule when combining all belief functions in cluster χ_a, c_0 is a domain dependent conflict, which is set to zero in this paper.

Minimizing the metaconflict function was the method of partitioning the belief functions into subsets representing the different events.

However, instead of considering the conflict in each cluster it is possible to refine the analysis and consider all pairwise conflicts between the belief functions in χ_a, $m_{ij}^{-}(\cdot) = c_{ij}$, where c_{ij} is the conflict of Dempster's rule when combining e_i and e_j. When $c_{ij} = 1$, e_i and e_j must not be in the same cluster, when $c_{ij} = 0$ there simply is no indication of the repellent type. It was demonstrated in [12] that minimizing a sum of logarithmized pairwise conflicts,

$$\sum_{a} \sum_{(kl)|e_k \wedge e_l \in \chi_a} -\log(1 - c_{kl}), \tag{4}$$

is a close approximation to minimizing the overall conflict, Eq. (3), making it possible to map the optimization problem onto a neural network for neural optimization.

3.1. Internal cluster conflicts and attractions

In addition to the conflicting metalevel evidence induced by the internal conflict between belief functions belonging to the same cluster, in many applications it is important to be able to handle attracting metalevel evidence from some external source stating that two belief functions concern the same object, Fig. 1. The mathematics of this problem was analyzed in [15].

Such an external metalevel evidence is represented as a pairwise piece of evidence $m_{ij}^{+}(\cdot) = p_{ij}$, where p_{ij} is a degree of attraction. When $p_{ij} = 1$, e_i and e_j must be in the same cluster, when $p_{ij} = 0$ we have no indication of the attracting type. However, let us first study the conflicts and their representation before we state the corresponding proposition of $m^+_{ij}(\cdot)$.

The frame for each cluster χ_a is refined as $\Theta_a = \{\text{AdP}, \neg\,\text{AdP}\} = \{\forall j.e_j \in \chi_a\}_1 \cup \{e_j \notin \chi_a\}_{j=1}^{|\chi_a|}$ where "adequate partition" AdP is equal to the singleton proposition $\forall j.e_j \in \chi_a$, that each belief function e_j placed in cluster χ_a actually belongs to χ_a. On the other hand, "not adequate partition" $\neg$AdP is refined to a set of $|\chi_a|$ propositions $e_j \notin \chi_a$, each stating that a particular belief function is misplaced. Representing the conflicting metalevel evidence as

$$m^-_{ij}(e_i \vee e_j \notin \chi_a) = c_{ij}, \quad m^-_{ij}(\Theta) = 1 - c_{ij}, \tag{5}$$

all conflicting metalevel evidence is combined within each cluster $\forall i,j,a.\ \oplus\{m^-_{ij}(\cdot)|e_i \wedge e_j \in \chi_a\}$ to derive

$$m^-_{\{\chi_a\}}(\neg\text{AdP}) = 1 - \prod_a \prod_{(ij)|e_i \wedge e_j \in \chi_a} (1 - c_{ij}),$$
$$m^-_{\{\chi_a\}}(\Theta) = 1 - m^-_{\chi}(\neg\text{AdP}). \tag{6}$$

Representing the attracting metalevel evidence as

$$m^+_{ij}(e_i \wedge e_j \in \chi_a) = p_{ij}, \quad m^+_{ij}(\Theta) = 1 - p_{ij}, \tag{7}$$

all attracting metalevel evidence is combined within each cluster $\forall i,j,a.\ \oplus\{m^+_{ij}(\cdot)|e_i \wedge e_j \in \chi_a\}$ to derive

$$m^+_{\{\chi_a\}}(\text{AdP}) = \prod_a \sum_{I \subseteq P_{|\chi_a|}|M_I \equiv N_{|\chi_a|}} \prod_I p_{ij} \prod_{P_{|\chi_a|}-I} (1 - p_{ij}),$$
$$m^+_{\{\chi_a\}}(\Theta) = 1 - m^+_{\chi}(\text{AdP}), \tag{8}$$

where $P_{|\chi_a|} = \{(ij)|1 \leqslant i < j \leqslant |\chi_a|\}$ is the set of all pairs of ordered numbers $\leqslant |\chi_a|$, $M_I = \{i|\exists p.(ip) \vee (pi) \in I\}$ is the set of all numbers in the pairs of I, and $N_{|\chi_a|} = \{1, \ldots, |\chi_a|\}$ is the set of all numbers $\leqslant |\chi_a|$.

This takes account of all internal cluster conflicts and all attractions between belief functions within each cluster.

3.2. External cluster conflicts and attractions

In this paper we supplement the above analysis by also taking into account any external conflicts and attractions between belief functions that are placed in different clusters.

While for the internal metalevel evidence we wanted to find a partition that minimized all conflicts and maximized all attractions, the situation is reversed for the external metalevel evidence. When belief functions are placed in different clusters, we like to see a maximization of conflicts and a minimization of attractions between these belief functions.

We combine all conflicting metalevel evidence for belief functions that are in different clusters $\forall i,j,a.$ $\oplus\{m^-_{ij}(\cdot)|e_i \wedge e_j \notin \chi_a\}$. From the result of this combination we derive

$$m^-_{\chi}(\text{AdP}) = \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} m^-_{ij}(e_i \vee e_j \notin \chi_a) = \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} c_{ij},$$
$$m^-_{\chi}(\Theta) = 1 - m^-_{\chi}(\text{AdP}). \tag{9}$$

In addition we also combine all attracting metalevel evidence for belief functions that are in different clusters $\forall i,j,a.\ \oplus\{m^+_{ij}(\cdot)|e_i \wedge e_j \notin \chi_a\}$. From the result of this combination we derive

$$m^+_{\chi}(\neg\text{AdP}) = 1 - \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} [1 - m^+_{ij}(e_i \wedge e_j \in \chi_a)]$$
$$= 1 - \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} (1 - p_{ij}),$$
$$m^+_{\chi}(\Theta) = 1 - m^+_{\chi}(\neg\text{AdP}). \tag{10}$$

3.3. Combine internal and external evidence

In Section 3.1 we derived $m^-_{\{\chi_a\}}(\neg\text{AdP})$ and $m^+_{\{\chi_a\}}(\text{AdP})$ based on all conflicting and all attracting internal metalevel evidence, respectively. In Section 3.2 we derived $m^-_{\chi}(\text{AdP})$ and $m^+_{\chi}(\neg\text{AdP})$ based on all conflicting and all attracting external metalevel evidence, respectively. We will here separately combine all evidence against and in favor of the partition.

We combine all evidence against the adequacy of the partition, $m^-_{\{\chi_a\}}(\neg\text{AdP})$ and $m^+_{\chi}(\neg\text{AdP})$, Eqs. (6) and (10), to receive

$$m^{-+}_{\{\chi_a\}\oplus\chi}(\neg\text{AdP}) = 1 - \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} (1 - p_{ij}) \times \prod_a \prod_{(ij)|e_i \wedge e_j \in \chi_a} (1 - c_{ij}),$$
$$m^{-+}_{\{\chi_a\}\oplus\chi}(\Theta) = 1 - m^{-+}_{\{\chi_a\}\oplus\chi}(\neg\text{AdP}), \tag{11}$$

and combine all evidence in favor of the adequacy of the partition, $m^+_{\{\chi_a\}}(\text{AdP})$ and, $m^-_{\chi}(\text{AdP})$, Eqs. (8) and (9), to receive

$$m^{+-}_{\{\chi_a\}\oplus\chi}(\text{AdP}) = 1 - \left(1 - \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} c_{ij}\right) \times \left[1 - \prod_a \sum_{I \subseteq P_{|\chi_a|}|M_I \equiv N_{|\chi_a|}} \prod_I p_{ij} \prod_{P_{|\chi_a|}-I} (1 - p_{ij})\right],$$
$$m^{+-}_{\{\chi_a\}\oplus\chi}(\Theta) = 1 - m^{+-}_{\{\chi_a\}\oplus_\chi}(\text{AdP}). \tag{12}$$

As the final step we combine all evidence against and in favor of the partition, Eqs. (11) and (12). We receive

$$m_{\{\chi_a\}\oplus\chi}(\text{AdP}) = [1 - m^{-+}_{\{\chi_a\}\oplus\chi}(\neg\text{AdP})]m^{+-}_{\{\chi_a\}\oplus\chi}(\text{AdP}),$$
$$m_{\{\chi_a\}\oplus\chi}(\neg\text{AdP}) = m^{-+}_{\{\chi_a\}\oplus\chi}(\neg\text{AdP})[1 - m^{+-}_{\{\chi_a\}\oplus\chi}(\text{AdP})],$$
$$m_{\{\chi_a\}\oplus\chi}(\Theta) = [1 - m^{-+}_{\{\chi_a\}\oplus\chi}(\neg\text{AdP})][1 - m^{+-}_{\{\chi_a\}\oplus\chi}(\text{AdP})],$$
$$m_{\{\chi_a\}\oplus\chi}(\emptyset) = m^{-+}_{\{\chi_a\}\oplus\chi}(\neg\text{AdP})m^{+-}_{\{\chi_a\}} \oplus_\chi (\text{AdP}). \tag{13}$$

This is the amount of support awarded to the proposition that we have an "adequate partition" $m_{\{\chi_a\}\oplus\chi}(\mathrm{AdP})$, and awarded to the proposition that we do not have an "adequate partition", $m_{\{\chi_a\}\oplus\chi}(\neg\mathrm{AdP})$ respectively, when taking everything into account. [1]

In Section 5 we will investigate how to minimize $m_{\{\chi_a\}\oplus\chi}(\neg\mathrm{AdP})$ by finding the best partition of evidence.

4. Potts spin theory

The Potts spin problem consists of minimizing an energy function

$$E = \frac{1}{2}\sum_{i,j=1}^{N}\sum_{a=1}^{q}(J_{ij}^{-} - J_{ij}^{+})S_{ia}S_{ja} \tag{14}$$

by changing the states of the spins S_{ia}'s, where $S_{ia} \in \{0, 1\}$ and $S_{ia} = 1$ means that report i is in cluster a. This model serves as a clustering method if J_{ij}^{-} is used as a penalty factor when report i and j are in the same cluster, and J_{ij}^{+} when they are in different clusters.

The minimization is carried out by simulated annealing. In simulated annealing temperature is an important parameter. The process starts at a high temperature where the S_{ia} change state more or less at random taking little account of the interactions (J_{ij}'s). The process continues by gradually lowering the temperature. As the temperature is lowered the random flipping of spins gradually come to a halt and the spins gradually become more influenced by the interactions (J_{ij}'s) so that a minimum of the energy function, Eq. (14), is reached. This gives us the best partition of all evidence into the clusters with minimal overall conflict.

For computational reasons we use a mean field model, where spins are deterministic with $V_{ia} = \langle S_{ia}\rangle$ (thermal averages), $V_{ia} \in [0, 1]$, in order to find the minimum of the energy function. The Potts mean field equations are formulated [10] as

$$V_{ia} = \frac{\mathrm{e}^{-H_{ia}[V]/T}}{\sum_{b=1}^{K}\mathrm{e}^{-H_{ib}[V]/T}}, \tag{15}$$

where

$$H_{ia}[V] = \sum_{j=1}^{N} J_{ij}V_{ja} - \gamma V_{ia} + \alpha\sum_{j=1}^{N} V_{ja}. \tag{16}$$

INITIALIZE

K (the problem size); $N = 2^K - 1$;

$J_{ij}^{-} = -\log(1 - c_{ij})\delta_{|A_i \cap A_j|} \quad \forall i, j$;

$J_{ij}^{+} = -\log(1 - p_{ij})(1 - \delta_{|A_i \cap A_j|}) \quad \forall i, j$;

$s = 0$; $t = 0$; $\varepsilon = 0.001$; $\tau = 0.9$; α (for $K \le 7$: $\alpha = 0$, $K = 8$: $\alpha = 10^{-6}$, $K = 9$: $\alpha = 0$, $K = 10$: $\alpha = 3 \cdot 10^{-7}$, $K = 11$: $\alpha = 3 \cdot 10^{-8}$); $\gamma = 0.5$;

$T^0 = T_c$ (a critical temperature) $= \frac{1}{K} \cdot max(-\lambda_{min}, \lambda_{max})$,

where λ_{min} and λ_{max} are the extreme eigenvalues of M,

where $M_{ij} = J_{ij}^{-} - J_{ij}^{+} + \alpha - \gamma\delta_{ij}$;

$V_{ia}^{0} = \frac{1}{K} + \varepsilon \cdot rand[0,1] \quad \forall i, a$;

REPEAT

- REPEAT–2

  $\forall i$ Do:

  - $H_{ia}^{s} = \sum_{j=1}^{N}(J_{ij}^{-} - J_{ij}^{+} + \alpha)V_{ja}^{\begin{cases} s+1, & j<i \\ s, & j \ge i \end{cases}} - \gamma V_{ia}^{s} \quad \forall a$;
  - $F_{i}^{s} = \sum_{a=1}^{K} e^{-H_{ia}^{s}/T^{t}}$;
  - $V_{ia}^{s+1} = \frac{e^{-H_{ia}^{s}/T^{t}}}{F_{i}^{s}} + \varepsilon \cdot rand[0,1] \quad \forall a$;
  - $s = s + 1$;

  UNTIL–2

  $\frac{1}{N}\sum_{i,a}\left|V_{ia}^{s} - V_{ia}^{s-1}\right| \le 0.01$;

- $T^{t+1} = \tau \cdot T^{t}$;
- $t = t + 1$;

UNTIL

$\frac{1}{N}\sum_{i,a}(V_{ia}^{s})^2 \ge 0.99$;

RETURN

$\left\{\chi_a \middle| \forall S_i \in \chi_a.\ \forall b \ne a\ V_{ia}^{s} > V_{ib}^{s}\right\}$;

Fig. 2. Clustering algorithm.

In order to minimize the energy function Eqs. (15) and (16) are iterated until a stationary equilibrium state has been reached for each temperature. Then, the temperature is lowered step by step by a constant factor until $\forall i, a.V_{ia} = 0, 1$ in the stationary equilibrium state, Fig. 2.

5. Mapping a Dempster–Shafer clustering problem onto potts spin

In order to map the function $m_{\{\chi_a\}\oplus\chi}(\neg\mathrm{AdP})$ onto a Potts spin neural network we must make one approximation and rewrite the function as a sum of terms that is to be minimized. For large scale problems we have few pairwise conflicts. If two different belief functions are drawn randomly we have (in our test case, Section 6) a probability of conflict between them of

$$P(A_i \cap A_j = \varnothing) = \frac{\sum_{j=1}^{K-1}\binom{K}{j}\sum_{k=1}^{K-j}\binom{K-j}{k}}{(2^K - 1)^2 - (2^K - 1)}, \tag{17}$$

[1] The observant reader will notice that we have not normalized in Eq. (13). While this approach is in spirit equal to Smets' transferable belief model (TBM) [17] it is done here simply for simplicity. Since we will only use $m_{\{\chi_a\}\oplus\chi}(\neg\mathrm{AdP})$ as a measure in optimization, its absolute value is without significance. Simply put, the question of normalizing or not is outside of the scope of our method. Thus, we consider the partitioning method described in this and previous articles to be compatible with both Shafer's and Smets' approaches.

where A_i and A_j are focal elements, K is the number of clusters and, e.g., $P(A_i \cap A_j = \emptyset) = 3.12\%$ when $K = 12$. When the information content is that low the product term of conflicts may be approximated by zero. If the attracting metalevel evidence has a similar low information content then product terms of attractions will also be approximated by zero. [2] Remaining will be all terms from probabilistic sums like

$$\prod_{(ij)} (1 - c_{ij}), \tag{18}$$

and

$$\prod_{(ij)} (1 - p_{ij}). \tag{19}$$

We make this low information content assumption and the corresponding approximation to $m_{\{\chi_a\}\oplus\chi}(\neg\text{AdP})$ in Eq. (13);

$$\begin{aligned} & m_{\{\chi_a\}\oplus\chi}(\neg\text{AdP}) \\ &= \left[1 - \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} (1 - p_{ij}) \times \prod_a \prod_{(ij)|e_i \wedge e_j \in \chi_a} (1 - c_{ij})\right] \\ &\quad \times \left(1 - \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} c_{ij}\right) \\ &\quad \times \left[1 - \prod_a \sum_{I \subseteq P_{|\chi_a|}|M_I \equiv N_{|\chi_a|}} \prod_I p_{ij} \prod_{P_{|\chi_a|}-I} (1 - p_{ij})\right] \\ &\approx \left[1 - \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} (1 - p_{ij}) \times \prod_a \prod_{(ij)|e_i \wedge e_j \in \chi_a} (1 - c_{ij})\right], \end{aligned} \tag{20}$$

where the last approximation in Eq. (20) is equal to Eq. (11).

Note that, in the last term before the approximation, the minimum cardinality of I is $\lceil |\chi_a|/2 \rceil$. Thus, the probability that

$$\prod_I p_{ij} = 0 \tag{21}$$

given our assumption is greater or equal to

$$1 - P(A_i \cap A_j = \emptyset)^{\lceil |\chi_a|/2 \rceil}. \tag{22}$$

In order to map the approximated $m_{\{\chi_a\}\oplus\chi}(\neg\text{AdP})$ onto a Potts spin neural network we need to rewrite Eq. (20) as a sum of terms. Let us rewrite the minimization as follows:

$$\begin{aligned} & \min m_{\{\chi_a\}\oplus\chi}(\neg\text{AdP}) \\ &\quad \approx \min 1 - \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} (1 - p_{ij}) \cdot \prod_a \prod_{(ij)|e_i \wedge e_j \in \chi_a} (1 - c_{ij}) \\ &\quad \Longleftrightarrow \max \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} (1 - p_{ij}) \cdot \prod_a \prod_{(ij)|e_i \wedge e_j \in \chi_a} (1 - c_{ij}) \\ &\quad \Longleftrightarrow \max \log \prod_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} (1 - p_{ij}) \cdot \prod_a \prod_{(ij)|e_i \wedge e_j \in \chi_a} (1 - c_{ij}) \\ &\quad = \max \sum_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} \log(1 - p_{ij}) \\ &\qquad + \sum_a \sum_{(ij)|e_i \wedge e_j \in \chi_a} \log(1 - c_{ij}) \\ &\quad = \min \sum_{(ij)|\forall a.e_i \wedge e_j \notin \chi_a} -\log(1 - p_{ij}) \\ &\qquad + \sum_a \sum_{(ij)|e_i \wedge e_j \in \chi_a} -\log(1 - c_{ij}). \end{aligned} \tag{23}$$

In order to apply the Potts model to Dempster–Shafer clustering we use interactions $J_{ij}^- = -\log(1 - c_{ij})\delta_{|A_i \cap A_j|}$ and $J_{ij}^+ = -\log(1 - p_{ij})(1 - \delta_{|A_i \cap A_j|})$ in the energy function Eq. (14), where δ is the Kronecker delta with

$$\delta_{|A_i \cap A_j|} \equiv \begin{cases} 1, & A_i \cap A_j = \emptyset, \\ 0, & \text{otherwise}, \end{cases} \tag{24}$$

where A_i and A_j are two focal elements and

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \tag{25}$$

in Fig. 2.

By minimizing the energy function we also minimize $m_{\{\chi_a\}\oplus\chi}(\neg\text{AdP})$. In Fig. 2 an algorithm for minimizing the energy function through iteration of Eqs. (15) and (16) is shown.

6. Results

In this section we compare the clustering performance and computational complexity of two different clustering methods. First, Potts spin clustering using conflicting evidence only. Secondly, Potts spin using both conflicting and attracting evidence. For each method, and all problem sizes, we will cluster $2^K - 1$ belief functions into K clusters. The frame of discernment is the set of natural numbers between 1 and K, $\Theta = \{1, 2, 3, \ldots, K\}$. Each of the $2^K - 1$ subsets of the frame receives support from exactly one belief function. Thus, we only use simple support functions in this test, but this can easily be extended to consonant belief functions.

[2] It is a reasonable assumption that this is the case in many real-world situations that have a high data rate. With a high data rate many pairwise conflicts will tend to be small or equal to zero as physical constraints resulting in conflicts are usually not violated in applications. For large scale problems this product term will fast approach zero as the problem size increases. The same argument can be made for the product terms of attractions, but to a somewhat lesser degree. Although it will certainly be possible to construct special situations when these product terms may not be approximated with zero, such situations would tend to have small data sets making them uninteresting for this method. In a real-world situation a previous version [2] of the method was implemented successfully using conflicts only [4].

Each belief function is given a mass number randomly selected in [0,1]. Although K $(=|\Theta|$; the number of clusters) and N $(=|2^{\Theta}|-1$; the number of belief functions) are not changed independently in the test examples, evidence is rather striking (Table 1) that the Potts spin computation time scales as $N^2\log^2 N$. A small overhead is noted for the smallest problem sizes.

In Table 2 we notice the clustering performance when using only conflicting evidence. A near perfect clustering performance is demonstrated for this problem with a near constant mean metaconflict per belief function for a few orders of magnitude. This demonstrates the adequacy of using a Potts spin neural network for Dempster–Shafer clustering.

The metaconflict (Mcf) in Table 2 is defined in Eq. (3), [11]. The metaconflict per cluster $\langle c_i\rangle$ was defined in [12] as

$$\langle c_i\rangle = 1-(1-\mathrm{Mcf})^{1/K}, \tag{26}$$

and the metaconflict per belief function as $\langle c_i\rangle K/N$.

In order to make a comparison between using conflicting only with using both conflicting and attracting metalevel evidence we need another measure that is objective to both methods. We use target classification error rate as the measure where a target identity is randomly selected from the focal element of each belief function. This becomes a much harder clustering problem. While two belief functions with subsets $\{1,2\}$ and $\{2,3\}$ will have no conflict when placed in the same cluster as they have a nonempty intersection this will be seen as a classification error unless both belief functions have number 2 selected as their target.

In this test we set $p_{ij}=rand[0,1)$ with a frequency such that $|\{p_{ij}|p_{ij}\neq 0\}|=|\{c_{ij}|c_{ij}\neq 0\}|$, and $p_{ij}=0$ for the remainder. Thus, we let the attracting and conflicting metalevel evidence be of equal importance.

In Table 3 we notice an improvement in the classification error rate of 76–80% (in the median) for some of the larger problem sizes when including attracting metalevel evidence compared to when using conflicting metalevel evidence only. This is achieved while also receiving a small reduction in computation time for most problem sizes, Table 1.

In Fig. 3 we observe the convergence of two clustering processes. Each line is a path traveled by a belief function from the center of the circle at the first

Table 1
Computation time using conflicting evidence only or both conflicting and attracting evidence simultaneously. Each number is the mean of ten randomly generated problems. All times are measured in seconds, running CMU Common Lisp 18d+ using MatLisp 1.0b on an Intel Pentium 4 (2.667 GHz CPU, 1.024 GB RAM) with Linux

No. of clusters, K	No. of items of evidence, N	Potts spin using conflicting evidence only		Potts spin using conflicting and attracting evidence	
		Time	Time/N^2K^2	Time	Time/N^2K^2
3	7	0.023	5.22×10^{-5}	0.015	3.40×10^{-5}
4	15	0.212	5.89×10^{-5}	0.101	2.81×10^{-5}
5	31	0.120	5.00×10^{-6}	0.101	4.20×10^{-6}
6	63	0.338	2.37×10^{-6}	0.247	1.73×10^{-6}
7	127	0.906	1.15×10^{-6}	0.988	1.25×10^{-6}
8	255	9.59	2.31×10^{-6}	2.02	4.85×10^{-7}
9	511	11.0	5.21×10^{-7}	8.57	4.05×10^{-7}
10	1023	56.9	5.44×10^{-7}	67.7	6.47×10^{-7}
11	2047	484	9.54×10^{-7}	353	6.97×10^{-7}
12	4095	3540	1.46×10^{-7}	3520	1.46×10^{-6}

Table 2
Metaconflict, metaconflict per cluster and metaconflict per belief function (median and mean over ten runs) using conflicting evidence only

No. of clusters, K	No. of items of evidence, N	Metaconflict			Metaconflict per cluster		Metaconflict per belief function	
		Median	Mean	Standard deviation	Median	Mean	Median	Mean
3	7	0	0.001	0.006	0	0.0004	0	0.0002
4	15	0	0.009	0.031	0	0.002	0	0.0006
5	31	0	0.023	0.098	0	0.005	0	0.0008
6	63	0	0.030	0.074	0	0.005	0	0.0005
7	127	0.001	0.029	0.070	0.0002	0.004	0.00001	0.0002
8	255	0.004	0.040	0.080	0.0005	0.005	0.00002	0.0002
9	511	0.177	0.248	0.289	0.021	0.031	0.0004	0.0005
10	1023	0.205	0.331	0.361	0.023	0.039	0.0002	0.0004
11	2047	0.877	0.728	0.306	0.173	0.112	0.0009	0.0006
12	4095	0.999	0.885	0.263	0.524	0.165	0.002	0.0005

Table 3
Classification error rate (median and mean over ten runs) using conflicting evidence only or both conflicting and attracting evidence simultaneously

No. of clusters, K	No. of items of evidence, N	Potts spin using conflicting evidence only			Potts spin using conflicting and attracting evidence		
		Median	Mean	Standard deviation	Median	Mean	Standard deviation
3	7	0.214	0.257	0.227	0	0.100	0.235
4	15	0.467	0.420	0.189	0	0.073	0.179
5	31	0.468	0.497	0.128	0.113	0.123	0.201
6	63	0.595	0.578	0.065	0.063	0.121	0.176
7	127	0.638	0.652	0.056	0.118	0.157	0.228
8	255	0.694	0.700	0.024	0.218	0.236	0.207
9	511	0.740	0.741	0.020	0.150	0.201	0.204
10	1023	0.764	0.762	0.008	0.181	0.218	0.196
11	2047	0.795	0.794	0.007	0.174	0.239	0.195
12	4095	0.815	0.813	0.005	0.163	0.247	0.190

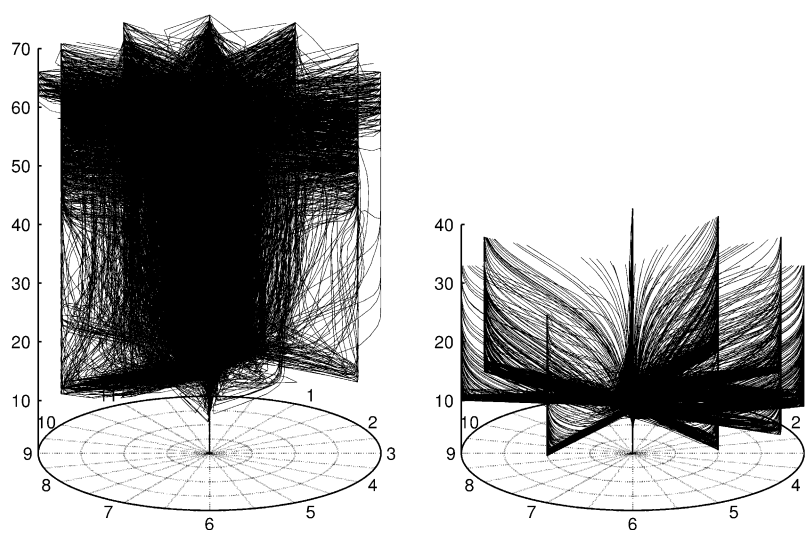

Fig. 3. The clustering process of 4095 belief function into twelve clusters. Left using conflicts only in 66 successive iterations (vertical axis). Right using attractions and conflicts in 33 iterations.

iteration of the neural network towards one of the twelve cluster positions at the edge of the circle. As the neural network gradually converges from an initial starting position with a $1/K$ output signal for each neuron to a situation with a 1 output signal for one neuron (corresponding to the cluster where the belief function is placed) and a 0 output signal for all other neurons, these output signals can be interpreted as partial memberships towards the different clusters. In Fig. 3 the twelve clusters are placed on a circle with radius 1. Each output signal of a neuron is represented by a vector from the center of the circle pointing towards the corresponding cluster position at the edge of the circle. The vector is scaled by the absolute value of the output signal. Each belief function is plotted as a point in the cylindrical plane of the figure as the weighted average of the twelve vectors and by iteration step on the vertical axis. Thus, a position along the path (starting from the center of the circle at iteration step 1 and terminating at one of the twelve cluster positions at the final iteration step) is a visualization of the weighted average of all partial cluster memberships for the belief function during the convergence. The figure on the left shows convergence when using conflicts only. Here the convergence is only gradual with much of the convergence taking place in the last dozen iterations. Right we observe the process when using both conflicts and attractions. Now much of the convergence takes place between iteration 10–12 with a few latecomers gradually converging thereafter.

In Fig. 4 we see a top view of the two processes. Notice in the left part of the figure that when using conflicts only many belief functions have to change course from one cluster to another as the process converges. When using both conflicts and attractions most belief function head directly for an appropriate cluster, right part of the figure.

Finally in Fig. 5 we observe the convergence of the two processes. Each curve is a normalized entropy-like measure

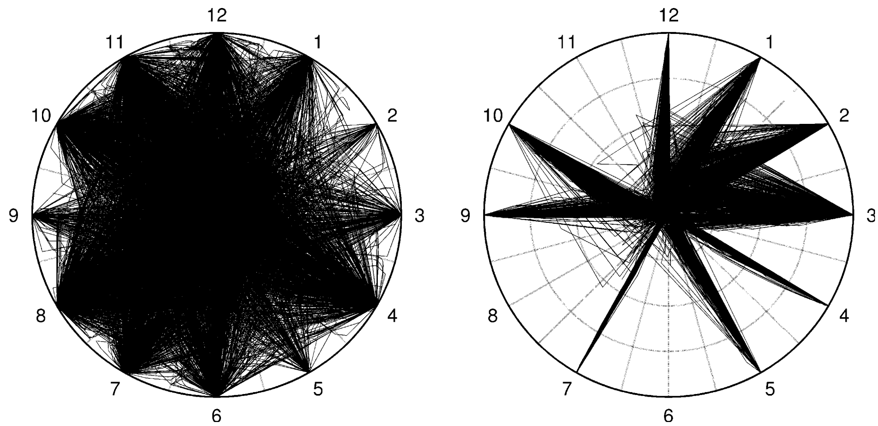

Fig. 4. A top view of Fig. 3. Left using conflicts only and right using attractions and conflicts.

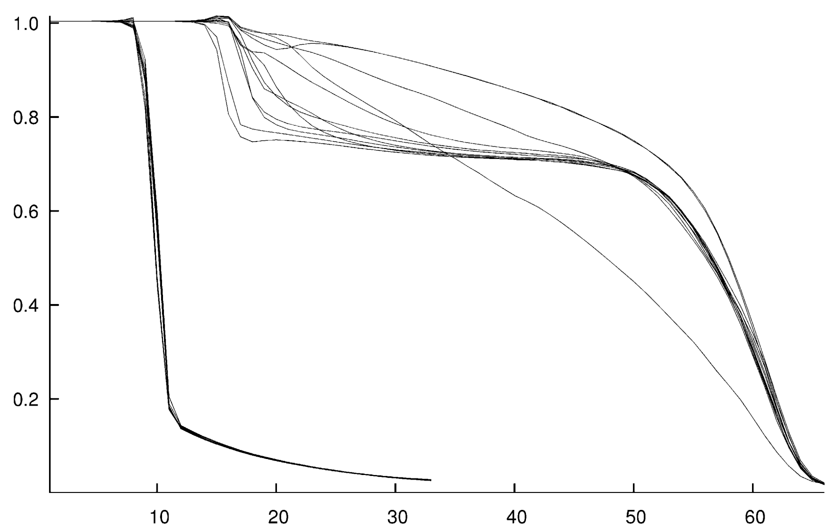

Fig. 5. The convergence of two cluster processes. Each curve is a normalized entropy-like measure of the neural output of the 4095 neurons corresponding to the cluster.

$$\sum_{ia} V_{ia} \log_2(V_{ia}) \qquad (27)$$

of each cluster. To the left, twelve curves illustrate the rapid convergence in 33 iterations of the twelve clusters when using both conflicts and attractions. To the right, twelve curves illustrate the more gradual convergence in 66 iterations when using conflicts only.

7. Conclusions

We have shown how to map a Dempster–Shafer clustering problem using both conflicting and attracting metalevel evidence onto a Potts spin neural network in order to obtain a fast and accurate method for large scale clustering problems.

Compared to the situation when using only conflicting metalevel evidence the new clustering method offers a significant reduction in classification errors as well as a small reduction in computation time.

References

[1] A. Ayoun, P. Smets, Data association in multitarget detection using the transferable belief model, International Journal of Intelligent Systems 16 (10) (2001) 1167–1182.
[2] M. Bengtsson, J. Schubert, Dempster–Shafer clustering using potts spin mean field theory, Soft Computing 5 (3) (2001) 215–228.
[3] S.S. Blackman, Association and fusion of multiple sensor data, in: Y. Bar-Shalom (Ed.), Multitarget–Multisensor Tracking: Advanced Applications, Artech House, Norwood, MA, 1990, pp. 187–218, Chapter 6.
[4] J. Cantwell, J. Schubert, J. Walter, Conflict-based force aggregation, in: Cd Proceedings of the Sixth International Command and Control Research and Technology Symposium, Annapolis, USA, 19–21 June 2001, US Department of Defence CCRP, Washington, DC, Track 7, Paper 031, pp. 1–15.
[5] A.P. Dempster, A generalization of Bayesian inference, Journal of the Royal Statistical Society B 30 (2) (1968) 205–247.

[6] T. Denœux, M. Masson, Clustering of proximity data using belief functions, in: B. Bouchon-Meunier, L. Foulloy, R.R. Yager (Eds.), Intelligent Systems for Information Processing: From Representation to Applications, Elsevier, Amsterdam, 2003, pp. 291–302.

[7] M. Grabisch, H. Prade, The correlation problem in sensor fusion in a possibilistic framework, International Journal of Intelligent Systems 16 (11) (2001) 1273–1283.

[8] Informations system Flygvapnet version 1.0 (IS FV II 1.0A), Swedish Armed Forces, 2002.

[9] N. Milisavljevic, Analysis and fusion using belief function theory of multisensor data for close-range humanitarian mine detection, Ph.D. Thesis, ENST 2001 E 012, École Nationale Supérieure des Télécommunications, Paris, 2001.

[10] C. Peterson, B. Söderberg, A new method for mapping optimization problems onto neural networks, International Journal of Neural Systems 1 (1) (1989) 3–22.

[11] J. Schubert, On nonspecific evidence, International Journal of Intelligent Systems 8 (6) (1993) 711–725.

[12] J. Schubert, Fast Dempster–Shafer clustering using a neural network structure, in: B. Bouchon-Meunier, R.R. Yager, L.A. Zadeh (Eds.), Information, Uncertainty and Fusion, Kluwer Academic Publishers (SECS 516), Boston, MA, 1999, pp. 419–430.

[13] J. Schubert, Managing inconsistent intelligence, in: Proceedings of the Third International Conference on Information Fusion, Paris, France, 10–13 July 2000, International Society of Information Fusion, Sunnyvale, CA, pp. TuB4/10–16.

[14] J. Schubert, Robust report level cluster-to-track fusion, in: Proceedings of the Fifth International Conference on Information Fusion, Annapolis, USA, 8–11 July 2002, International Society of Information Fusion, Sunnyvale, CA, pp. 913–918.

[15] J. Schubert, Clustering belief functions based on attracting and conflicting metalevel evidence, in: B. Bouchon-Meunier, L. Foulloy, R.R. Yager (Eds.), Intelligent Systems for Information Processing: From Representation to Applications, Elsevier, Amsterdam, 2003, pp. 349–360.

[16] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.

[17] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (2) (1994) 191–234.

[18] P. Smets, Practical uses of belief functions, in: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, Stockholm, Sweden, 30 July–1 August 1999, Morgan Kaufmann Publishers, San Francisco, CA, pp. 612–621.

[19] P. Smets, Data fusion in the transferable belief model, in: Proceedings of the Third International Conference on Information Fusion, Paris, France, 10–13 July 2000, International Society of Information Fusion, Sunnyvale, CA, PS/21–33.

[20] F.Y. Wu, The Potts model, Reviews of Modern Physics 54 (1) (1982) 235–268.