Available online at www.sciencedirect.com
ScienceDirect
International Journal of Approximate Reasoning
48 (2008) 466–480
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

Clustering decomposed belief functions using generalized weights of conflict ☆

Johan Schubert *

Department of Decision Support Systems, Division of Command and Control Systems, Swedish Defence Research Agency, SE-164 90 Stockholm, Sweden

Received 30 June 2006; received in revised form 9 March 2007; accepted 20 March 2007
Available online 3 April 2007

Abstract

We develop a method for clustering all types of belief functions, in particular non-consonant belief functions. Such clustering is done when the belief functions concern multiple events, and all belief functions are mixed up. Clustering is performed by decomposing all belief functions into simple support and inverse simple support functions that are clustered based on their pairwise generalized weights of conflict, constrained by weights of attraction assigned to keep track of all decompositions. The generalized conflict $c \in (-\infty, \infty)$ and generalized weight of conflict $J^{-} \in (-\infty, \infty)$ are derived in the combination of simple support and inverse simple support functions.

Keywords: Dempster–Shafer theory; Decomposition; Clustering; Generalized weight of conflict; Simulated annealing; Inverse simple support functions; Belief function; Non-consonant belief function; Pseudo belief function

1. Introduction

In earlier papers [2–5] we developed methods within Dempster–Shafer theory [6–8] to manage simple support functions (SSFs) that concern different events where the SSFs were mixed up. This was the case when it was not known a priori to which event each SSF was related. The SSFs were clustered into subsets that should be handled independently. This was based on minimizing pairwise conflicts within each cluster where conflicts served as repellence, forcing conflicting SSFs into different clusters.

This method was extended [9,10] into also handling external information of an attracting nature, where attractions between SSFs suggested they belonged together.

☆ A short version of this study was presented at the Eleventh International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU 2006) in Paris, France [1].

* Tel.: +46 8 55503702; fax: +46 8 55503700.
E-mail address: schubert@foi.se
URL: http://www.foi.se/fusion.

0888-613X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2007.03.002

In this paper we develop a method for managing non-consonant belief functions concerning different events where the belief functions are mixed up.[1] This is the general case where no a priori information is available regarding which event the belief functions refer to. This method is based on the extension introducing attractions [9,10] and a decomposition method for belief functions [11].

In short, the method can be described as first decomposing all belief functions into a set of SSFs and inverse simple support functions (ISSFs) [11]. Secondly, all SSFs and ISSFs are clustered, taking account of both the conflicts between every pair of SSFs and ISSFs as well as information regarding which SSFs and ISSFs were decomposed from the same belief function.

The number of clusters in the clustering process is an input parameter that needs to be known a priori. However, determination of number of clusters is outside the scope of this paper. It can be managed with other methods, e.g., the sequential estimation method proposed by Schubert and Sidenbladh [12] or inferred through several different trials by the L-method [13].

The methodology developed in this paper is intended to manage intelligence reports whose uncertainty is represented as belief functions with several alternative nonspecific propositions, i.e., non-singleton focal elements. This can be the case when handling human intelligence (HUMINT) or for that matter sensor reports from some advanced type of sensor. For such sensors it is natural to think that the sensor resolution at different ranges will correspond to different sizes of the focal elements supported, while the internal representation of the frame of discernment will correspond to the maximum resolution. If such sensors can handle two or more alternative hypotheses as two or more focal elements that will be non-singletons except in the best of conditions at short ranges, this will force us to manage general belief functions. Presumably, humans as information sources will also on average deliver fewer but more complex intelligence reports than simple sensor systems. Such complex intelligence or advanced sensor reports can be decomposed and managed with these methods.

An earlier version of this method [2] is implemented in *Anubis*, a Swedish Army Intelligence System, and *ISFV*, a Swedish Air Force Intelligence System [14], and a later version [10] is implemented in *IFD03*, an information fusion demonstrator for tactical intelligence processing [15].

For a recent overview of different alternatives to manage the combination of conflicting belief functions, see the article by Smets [16].

We begin by describing the decomposition method for belief functions (Section 2). In Section 3 we study the characteristics of all types of combinations of SSFs and ISSFs and how generalized conflicts between SSFs and ISSFs are mapped onto weights. We demonstrate how to manage all SSFs and ISSFs using these weights together with logical constraints that keep track of the decomposition (Section 4). In Section 5 an example is given. Finally, in Section 6, conclusions are drawn.

2. Decomposition

Let us begin by defining an ISSF:

Definition. An inverse simple support function on a frame of discernment Θ is a function $m : 2^{\Theta} \rightarrow (-\infty, \infty)$ characterized by a weight $w \in (1, \infty)$ and a focal element $A \subseteq \Theta$, such that $m(\Theta) = w, m(A) = 1 - w$ and $m(X) = 0$ when $X \notin \{A, \Theta\}$.

Let us now recall the meaning of SSFs and ISSFs, [11]: An SSF $m_1(A) \in [0, 1]$ represents a state of belief that "You have some reason to believe that the actual world is in A (and nothing more)". An ISSF $m_2(A) \in (-\infty, 0)$ on the other hand, represents a state of belief that "You have some reason *not* to believe that the actual world is in A". Equivalently, in the terminology of [11], A_1^w where $w \in [0, 1]$ and A_2^w where $w \in (1, \infty)$, respectively. Here, w is the mass assigned to Θ in m_1 and m_2. Note the notation used, where A_1^w and A_2^w represent the support for identical subsets A of the frame given by two different SSFs or ISSFs with

[1] Consonant belief functions can be handled in the same way as SSFs without the method developed in this paper, by clustering the consonant belief functions without any decomposition using conflicts only [2].

index numbers 1 and 2. The lower index is the index number of the SSF or ISSF that lends support to this subset.

The ISSF A_2^w can be understood as some reason *not* to believe in A due to its absorbing belief. A simple example is one SSF $A_1^{3/4}$, i.e., $m_1(A) = 1/4$ and $m_1(\Theta) = 3/4$, and one ISSF $A_2^{4/3}$, i.e., $m_2(A) = -1/3$ and $m_2(\Theta) = 4/3$. Combining these two functions $A_1^{3/4} \oplus A_2^{4/3} = A_1^{3/4} \ominus A_2^{3/4} = A^1$ (i.e., $m_{1\oplus 2}(\Theta) = 1$) yields a vacuous belief function. Here, $\text{Bel}_x \oplus A_1^y = \text{Bel}_x \ominus A_1^{1/y}$, where $\oplus$ is Dempster's rule, and $\ominus$ is the decombination operator absorbing belief [11]. Thus, the ISSF $A_2^{4/3}$ can be interpreted as 1/4 reason *not* to believe in A, since it precisely eliminates the 1/4 support in A expressed by $A_1^{3/4}$. It means that if you previously had some 1/4 belief in A you should now delete it. That cannot be achieved by supporting the complement of A. This makes $A_1^{3/4}$ and $A_2^{4/3}$ into two unique components called *confidence* and *diffidence*, respectively, by Smets [11]. Now, if you start out with only one ISSF A^w, $w > 1$, and nothing more, this is interpreted as if you have *no* reason to believe in A and that you need more than $1/w$ additional reason before you will start believing in it. At precisely $1/w$ additional reason you will become completely ignorant $m(\Theta) = 1$. This is different than having some belief in A and some in A^c whose combination can never be reduced to complete ignorance.

All belief functions can be decomposed into a set of SSFs and ISSFs using the method developed by Smets [11]. The decomposition method is performed in two steps (Eqs. (1) and (2)). First, for any non-dogmatic belief function Bel_0, i.e., where $m_0(\Theta) > 0$, calculate the commonality number for all focal elements A by Eq. (1). We have

$$Q_0(A) = \sum_{B \supseteq A} m_0(B). \tag{1}$$

Secondly, calculate $m_i(C)$ for all decomposed SSFs and ISSFs, where $C \subseteq \Theta$ including $C = \emptyset$, and i is the ith SFF or ISSF. There will be one SSF or ISSF for each subset C of the frame unless $m_i(C)$ happens to be zero. In the general case we will have $|2^\Theta|$ SFFs and ISSFs. We get for all $C \subseteq \Theta$ including $C = \emptyset$

$$\begin{aligned} m_i(C) &= 1 - \prod_{A \supseteq C} Q_0(A)^{(-1)^{|A|-|C|+1}}, \\ m_i(\Theta) &= 1 - m_i(C). \end{aligned} \tag{2}$$

For dogmatic belief functions assign $m_0(\Theta) = \varepsilon > 0$ and discount all other focal elements proportionally.

For fast computation, take the logarithm of the product terms in Eq. (2) and use the fast Möbius transform [17].

Let us study a simple illustrative example of decomposing a general belief function Bel_0 into SSFs and ISSFs. Assume a frame $\Theta = \{x, y, z\}$ and random assignment for all focal elements,

$$\begin{aligned} m_0(\emptyset) &= 0 \\ m_0(\{x\}) &= 0.0505 \\ m_0(\{y\}) &= 0.1399 \\ m_0(\{x,y\}) &= 0.1176 \\ m_0(\{z\}) &= 0.1840 \\ m_0(\{x,z\}) &= 0.1877 \\ m_0(\{y,z\}) &= 0.2668 \\ m_0(\Theta) &= 0.0535. \end{aligned} \tag{3}$$

We calculate the commonality numbers for all subsets A of the frame, $A \subseteq \Theta$, using Eq. (1). We get,

$$
\begin{aligned}
&Q_0(\emptyset) = 1\\
&Q_0(\{x\}) = 0.4093\\
&Q_0(\{y\}) = 0.5777\\
&Q_0(\{x,y\}) = 0.1711\\
&Q_0(\{z\}) = 0.6920\\
&Q_0(\{x,z\}) = 0.2412\\
&Q_0(\{y,z\}) = 0.3202\\
&Q_0(\Theta) = 0.0535.
\end{aligned} \tag{4}
$$

From these commonality numbers we can calculate the support for each subset C of the frame, $C \subseteq \Theta$, for all seven SSFs and ISSFs excluding $m(\Theta)$ using Eq. (2). We get,

$$
\begin{aligned}
&m_1(\emptyset) = 0.3379 && m_1(\Theta) = 0.6621\\
&m_2(\{x\}) = -0.8856 && m_2(\Theta) = 1.8856\\
&m_3(\{y\}) = -0.7739 && m_3(\Theta) = 1.7739\\
&m_4(\{x,y\}) = 0.6875 && m_4(\Theta) = 0.3125\\
&m_5(\{z\}) = -1.0876 && m_5(\Theta) = 2.0876\\
&m_6(\{x,z\}) = 0.7783 && m_6(\Theta) = 0.2217\\
&m_7(\{y,z\}) = 0.8330 && m_7(\Theta) = 0.1670.
\end{aligned} \tag{5}
$$

We notice four SSF m_1, m_4, m_6 and m_7 and three ISSF m_2, m_3 and m_5. The three ISSF can be written $\{x\}_2^{1.8856}$, $\{y\}_3^{1.7739}$ and $\{z\}_5^{2.0876}$ using the terminology of [11].

3. Combining simple support functions and inverse simple support functions

When combining two decomposed parts from two different belief function we face three different situations: the combination of two SSFs, one SSF and one ISSF, or two ISSFs. These situations are studied below.

3.1. Two SSFs

In this situation we have two SSFs where $m_1(A) \in [0,1]$ and $m_2(B) \in [0,1]$. When the two simple support functions are combined we receive a conflict $c_{12} \in [0,1]$ whenever $A \cap B = \emptyset$. A weight of conflict is calculated by

$$J_{ij}^- = -\log(1 - c_{ij}) \tag{6}$$

where

$$c_{ij} = \begin{cases} m_i(A)m_j(B), & A \cap B = \emptyset \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

and $J_{ij}^- \in [0,\infty)$ but will be constrained to $J_{ij}^- \in [0,5]$ in our neural clustering process [2,10] for computational reasons. This will ensure convergence. The weight J_{ij}^- will work as repellence between m_i and m_j in the clustering process. We use the notation J_{ij}^- for a weight of conflict to differentiate it from J_{ij}^+, a weight of attraction that will be introduced in Section 4.

This is the usual situation. It is proper that two propositions referring to different conflicting hypotheses are not combined when they are highly conflicting. Using the conflict we obtain such a graded measure (see [3]).

3.2. One SSF and one ISSF

The situation when combining one SSF m_1 with one ISSF m_2 is interesting and unproblematic. Here, we have A_1^w where $w \in [0,1]$ as usual, and B_2^w where $w \in (1,\infty)$, i.e., in terms of mass functions $m_2(B) \in (-\infty, 0)$.

Thus, when we combine a SSF A_1^w with an ISSF B_2^w we receive a generalized conflict $c_{12} \in (-\infty, 0]$ whenever $A \cap B = \emptyset$. Using Eq. (6) we get a generalized weight of conflict $J_{12}^- \in (-\infty, 0]$ which will serve as a weak attraction between m_1 and m_2. As before we will constrain the generalized weight of conflict for computational reasons, here to $J_{ij}^- \in [-5, 0]$.

The weak attraction is proper and rather immediate. If you believe in a proposition A $(A_1^w, 0 \leqslant w \leqslant 1)$ and you receive further evidence indicating you have some reason *not* to believe in B $(B_1^w, w > 1)$, $A \cap B = \emptyset$, that is an indirect weak support of A as some alternatives of the frame not supported by m_1 are disbelieved.

A simple example will demonstrate this. Suppose you have an SSF $A_1^{1/2}$ and an ISSF $B_2^{3/2}$ such that $A \cap B = \emptyset$. Combining them will result in a new type of object, henceforth called a *pseudo belief function* [11].

In standard notation $A_1^{1/2}$ is

$$m_1(X) = \begin{cases} 1/2, & X = A \\ 1/2, & X = \Theta \end{cases} \tag{8}$$

and $B_2^{3/2}$ is

$$m_2(X) = \begin{cases} -1/2, & X = B \\ 3/2, & X = \Theta. \end{cases} \tag{9}$$

A straightforward combination of m_1 and m_2 yields a pseudo belief function

$$m_{1\oplus 2}(X) = \begin{cases} 3/4, & X = A \\ -1/4, & X = B \\ 3/4, & X = \Theta \\ -1/4, & X = \emptyset \end{cases} \tag{10}$$

without normalization and

$$m_{1\oplus 2}(X) = \begin{cases} 3/5, & X = A \\ -1/5, & X = B \\ 3/5, & X = \Theta \end{cases} \tag{11}$$

after normalization. This is an increase of m_1's support for A from 1/2 to 3/4 and 3/5, respectively, after combination with m_2. Note the interesting effect of normalization. Usually mass on the empty set is distributed proportionally among all focal elements by weighting up the support of the focal elements through normalization. When $m(\emptyset) < 0$, then instead the support for each focal element is weighted down to distribute support to the empty set so as to make $m(\emptyset) = 0$.

This support for the focal elements of is $m_{1\oplus 2}$ is different from the one we would have if we instead had received support for B^c of 1/2, $A \cap B = \emptyset$. Assume we have

$$m_3(X) = \begin{cases} 1/2, & X = B^c \\ 1/2, & X = \Theta, \end{cases} \tag{12}$$

then combining m_1 and m_3 yields

$$m_{1\oplus 3}(X) = \begin{cases} 1/2, & X = A \\ 1/4, & X = B^c \\ 1/4, & X = \Theta, \end{cases} \tag{13}$$

i.e., support for A of 1/2, or 3/4 if $B^c \equiv A$.

When two conflicting belief functions are decomposed, each into several SSFs and ISSFs, the total conflict for all pairs of two SSFs originating from different belief functions will be higher than that between the two belief functions. This is because the SSFs have higher masses on their focal elements than the corresponding belief function, now that we also have ISSFs with negative mass.

A simple example will demonstrate the situation. Let us assume two belief functions m_a and m_b whose basic belief assignments are

$$m_a(X) = \begin{cases} 1/2, & X = \{x, y\} \\ 3/10, & X = \{x, z\} \\ 1/5, & X = \Theta \end{cases} \tag{14}$$

and

$$m_b(X) = \begin{cases} 1/2, & X = \{x, y\} \\ 3/10, & X = \{y, q\} \\ 1/5, & X = \Theta. \end{cases} \tag{15}$$

The combination of m_a and m_b yields a conflict in the intersection of each function's second focal element $\{x, z\} \cap \{y, q\} = \emptyset$ of $m_{a\oplus b}(\emptyset) = 9/100$.

Using the decomposition algorithm, m_a can be decomposed into three functions. We get two SSFs $\{x, y\}_{a_1}^{2/7}$ and $\{x, z\}_{a_2}^{2/5}$, and one ISSF $\{x\}_{a_3}^{7/4}$, where $m_{a_1} \oplus m_{a_2} \oplus m_{a_3} = m_a$.

Similarly, m_b can be decomposed into two SSFs $\{x, y\}_{b_1}^{2/7}$ and $\{y, q\}_{b_2}^{2/5}$, and one ISSF $\{y\}_{b_3}^{7/4}$. Of the four pairs of SSFs (one from each decomposed belief function) only m_{a_2} and m_{b_2} are in conflict; $\{x, z\} \cap \{y, q\} = \emptyset$, see Fig. 1.

Combining m_{a_2} and m_{b_2} (or for that matter all four SSFs $m_{a_1}, m_{a_2}, m_{b_1}$, and m_{b_2}) yields a conflict $m_{a_2\oplus b_2}(\emptyset) = 9/25$, i.e., four times as much conflict as in the combination $m_a \oplus m_b$. This will be compensated for by a negative generalized conflict when including the two ISSFs m_{a_3} and m_{b_3} in the picture. We observe (in Fig. 1) generalized conflicts between m_{a_2} and m_{b_3}, and between m_{a_3} and m_{b_2}, respectively, i.e., $m_{a_2\oplus b_3}(\emptyset) = m_{a_3\oplus b_2}(\emptyset) = -9/20$.

3.3. Two ISSFs

The situation when combining two inverse simple support functions (ISSFs) m_1 and m_2 is perhaps the most interesting case. Here, we have two ISSFs A_1^w and B_2^w where $w \in (1, \infty)$, i.e., in terms of mass functions $m_1(A) \in (-\infty, 0)$ and $m_2(B) \in (-\infty, 0)$.

Assuming $A \cap B = \emptyset$, we receive a generalized conflict $c_{12} \in (0, \infty)$ when combining m_1 and m_2 that will serve as a repellence. This is proper but perhaps not immediately intuitive. Let us again look at an example. Let us combine $A_1^{3/2}$ and $B_2^{3/2}$, i.e., $m_1(A) = m_2(B) = -1/2$ or in other terms you have some (1/3) reason *not* to believe that the actual world is A and B, respectively, since $\text{Bel}_x \oplus A_1^{3/2} = \text{Bel}_x \ominus A_1^{2/3}$, where $\ominus$ is the decombination operator [11]. We have

$$m_1(X) = \begin{cases} -1/2, & X = A \\ 3/2, & X = \Theta \end{cases} \tag{16}$$

and

$$m_2(X) = \begin{cases} -1/2, & X = B \\ 3/2, & X = \Theta. \end{cases} \tag{17}$$

Combining m_1 and m_2 gives us

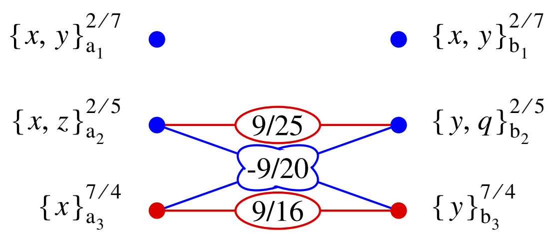

Fig. 1. Generalized conflicts between SSFs and ISSFs originating from m_a and m_b.

$$m_{1\oplus2}(X) = \begin{cases} -3/4, & X = A \\ -3/4, & X = B \\ 9/4, & X = \Theta \\ 1/4, & X = \emptyset \end{cases} \tag{18}$$

without normalization and

$$m_{1\oplus2}(X) = \begin{cases} -1, & X = A \\ -1, & X = B \\ 3, & X = \Theta \end{cases} \tag{19}$$

after normalization.

The positive conflict $c_{12} = 1/4$ will serve to repel m_1 and m_2 which is proper since m_1 and m_2 contradict each other. This is observed in the decrease of belief in $X = A$ and $X = B$ where $m_{1\oplus2}(A) < m_1(A)$ and $m_{1\oplus2}(B) < m_2(B)$, i.e., the reason to doubt that $X = A$ increases.

When the generalized conflict is greater than 1 we cannot use Eq. (6) to calculate a generalized weight of conflict as the logarithm is not defined for values less than 0. We call this *hyper conflicting*. We note, however, that the "1" in Eq. (6) is just a way to map a mass in the $[0, 1]$ interval to a weight in the $[0,\infty)$ interval. As there is nothing special about the "1" in Eq. (6) other than being an upper limit for a traditional conflict we can choose any other value greater than 1 to map hyper conflicts onto weights. One radical alternative would be to adjust the value to each application by choosing to map the interval $[0, \max\{c_{ij}|\forall i,j\}]$ to the interval $[0,\infty)$ in the case with two ISSFs or $(-\infty, \max\{c_{ij}|\forall i,j\})$ to $(-\infty,\infty)$ in the general case. We could redefine Eq. (6) as

$$J^-_{ij} = -\log(\max\{c_{kl}|\forall k,l\} - c_{ij}). \tag{20}$$

However, we will not do so. While this would work there are some drawbacks involved in choosing such a solution. Firstly, if the maximum value is very high compared to most other generalized conflicts, most generalized weights of conflict would be very small which would lead to a slow convergence in the clustering process. Secondly, having a generalized conflict mapped into different generalized weights of conflict depending on the application is not attractive. Thirdly, we would like to maintain consistency with clustering only SSFs where two SSFs that flatly contradict each other for a conflict of 1 also receive a weight of conflict of ∞ and nothing less.

Thus, we will map any hyper conflicting generalized conflict greater than one to a weight of ∞. For generalized conflicts less than 0 there are of course no problems. From this we may redefine Eq. (6) as

$$J^-_{ij} = -\log(1 - \min\{1, c_{ij}\}), \tag{21}$$

where $J^-_{ij} \in (-\infty,\infty)$. As before we will, however, for computational reasons restrict the generalized weight of conflict to $J^-_{ij} \in [-5, 5]$.

4. Clustering SSFs and ISSFs from decomposed belief functions

In order to be able to cluster all belief functions we begin by decomposing each belief function into a set of SSFs and ISSFs. We then calculate generalized weights of conflicts for all pairs in the set of decomposed SSFs and ISSFs, except when they both originate from the same belief function. Weights of attraction are assigned when they do originate from the same belief function.

At this stage all SSFs and ISSFs may now be clustered based on their pairwise generalized weights of conflict where the weights of attraction are used as constraints, forcing SSFs and ISSFs that originate from the same belief function to end up in the same cluster. This can be achieved using the Potts spin [18] neural clustering method extended with attractions [10].

The Potts spin problem consists of minimizing an energy function

$$E = \frac{1}{2}\sum_{i,j=1}^{N}\sum_{a=1}^{K}(J^-_{ij} - J^+_{ij})S_{ia}S_{ja} \tag{22}$$

by changing the states of the spins S_{ia}'s, where $S_{ia}\{0,1\}$ and $S_{ia}=1$ means that m_i is in cluster a. N is the number of SSFs and ISSFs, K is the number of clusters and $J^+_{ij} \in [0,\infty)$ is a weight of attraction formally calculated as

$$J^+_{ij} = -\log(1-p_{ij}), \tag{23}$$

where p_{ij} is a basic belief assignment that m_i and m_j originate from the same belief function. This model serves as a clustering method if J^-_{ij} is used as a penalty factor when m_i and m_j are in the same cluster.

However, if m_i and m_j originate from the same belief function we assign $c_{ij} := 0$ and an attraction $p_{ij} := 1$, otherwise $p_{ij} := 0$. To assure smooth convergence of the neural network J^-_{ij} is restricted to $[-5,5]$, while J^+_{ij} is restricted to $\{0,\beta\}$ in this application, $\beta > 0$.

Let us calculate the generalized weight of conflict between m_i and m_j, taking the restriction into account, Fig. 2 as

$$J^-_{ij} = \begin{cases} 0, & \exists x.m_i, m_j \in \mathrm{Bel}_x \\ -5, & \forall x.m_i, m_j \notin \mathrm{Bel}_x, \\ & c_{ij} \leqslant 1-e^5 \\ -\ln(1-c_{ij}), & \forall x.m_i, m_j \notin \mathrm{Bel}_x, \\ & 1-e^5 < c_{ij} < 1-e^{-5} \\ 5, & \forall x.m_i, m_j \notin \mathrm{Bel}_x, \\ & c_{ij} \geqslant 1-e^{-5}, \end{cases} \tag{24}$$

where $\exists x.m_i, m_j \in \mathrm{Bel}_x$ means there exists at least one belief function such that both m_i and m_j are decomposed from it, while $\forall x.m_i, m_j \notin \mathrm{Bel}_x$ means that there is no belief function such that both m_i and m_j are decomposed from it. We assign weights of attraction as

$$J^+_{ij} = \begin{cases} \beta, & \exists x.m_i, m_j \in \mathrm{Bel}_x \\ 0, & \forall x.m_i, m_j \notin \mathrm{Bel}_x \end{cases}, \tag{25}$$

where $\beta > 0$, enforcing the constraint that SSFs and ISSFs originating from the same belief function end up in the same cluster.

In the next section we will investigate suitable values for β in order to achieve a good balance between the repellency of the conflicts and the attractions from the constraints in the clustering process.

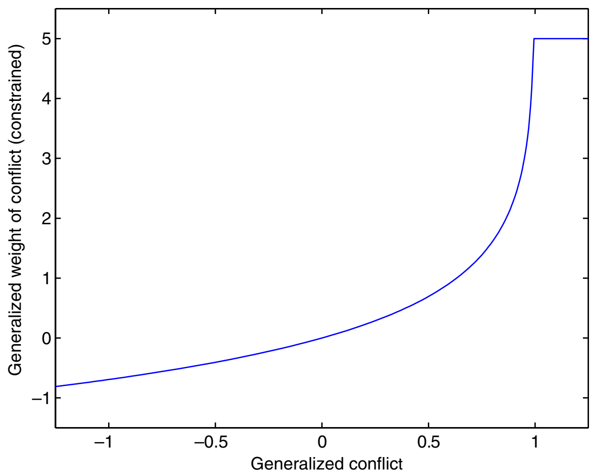

Fig. 2. The generalized weight of conflict J^-_{ij} in Eq. (24) as a function of the generalized conflict c_{ij} when $m_i, m_j \notin \mathrm{Bel}_x$.

The clustering of all SSFs and ISSFs is made using the Potts spin neural clustering method extended with attractions. The minimization of the energy function, Eq. (22), 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 comes to a halt and the spins gradually become more influenced by the interactions (J_{ij}'s) so that a minimum of the energy function 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 in order to find the minimum of the energy function. Here, spins are deterministic with $V_{ia} = \langle S_{ia} \rangle \in [0, 1]$ where V_{ia} is the expectation value of S_{ia}. The Potts mean field equations are formulated [19] as

$$V_{ia} = \frac{e^{-H_{ia}[V]/T}}{\sum_{b=1}^{K} e^{-H_{ib}[V]/T}} \tag{26}$$

where

$$H_{ia}[V] = \sum_{j=1}^{N} (J_{ij}^{-} - J_{ij}^{+}) V_{ja} - \gamma V_{ia} \tag{27}$$

and T is a temperature variable that is initialized to the critical temperature T_c, see Table 1, and then lowered step-by-step during the clustering process.

In order to minimize the energy function, Eqs. (26) and (27) 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 $\forall i,\ a.\ V_{ia} \in \{0, 1\}$ in the stationary equilibrium state, Table 1. We have two input parameters: K is the number of clusters and N the number of SSFs/ISSFs. Output is the set of all clusters $\{\chi_a\}$ where χ_a is cluster number a with all the SSFs and ISSFs that belong to it.

After clustering, SSFs and ISSFs originating from the same belief function may be replaced by the original belief function. The belief functions within each cluster can then be combined as a series of independent subproblems.

5. An example

In this section we will first go through a simple qualitative example to facilitate understanding of the mechanics of the clustering process. After this we will look at a larger quantitative experiment in order to assess the performance of the clustering process.

5.1. A qualitative discussion

Let us revisit the example in Section 3.2. We have two belief functions m_a and m_b, Eqs. (14) and (15), respectively. Let us duplicate m_a so that we now have three belief functions for this example (call the new belief function m_A). In Fig. 3, we have calculated the generalized weights of conflict J^- between all SSFs and ISSFs, and in Fig. 4 we have assigned the weights of attractions J^+ that will enforce the constraints. Only generalized weights of conflicts *not* equal to zero are included in Fig. 3, and only weights of attractions not equal to zero are included in Fig. 4.

The Potts spin neural clustering process will gradually decide which SSFs and ISSFs belong together. The convergence of the Potts spin neural net is fastest where the absolute value of the weight $|J^- - J^+|$ are largest. First, all SSFs and ISSF from the same belief function will be brought together due to the high weights from the constraints. Secondly, the conflict between the second SSF of m_b, $\{y, q\}_{b_2}^{2/5}$, towards the second SSFs of m_a and m_A, as well as that between the ISSF of m_b, $\{y\}_{b_3}^{7/4}$, towards the ISSFs of m_a and m_A will overwhelm the attraction between the SSFs and ISSFs, Fig. 3. Thus, the four SSFs and two ISSF from m_a and m_A are brought into one cluster and m_b will be left alone in a second cluster, assuming two clusters.

Table 1
Clustering algorithm

INITIALIZE
K (number of clusters); N (number of SSFs/ISSFs);
Calculate J^-_{ij} $\forall i, j$ using Eq. (24);
Assign J^+_{ij} $\forall i, j$ using Eq. (25);
$s = 0$; $t = 0$; $\varepsilon = 0.001$; $\tau = 0.9$; $\gamma = 0.5$;
$T^0 = T_c$ (a critical temperature)
$= \frac{1}{K} \cdot \max(-\lambda_{\min}, \lambda_{\max})$,
where $\lambda_{\min}$ and
$\lambda_{\max}$ are the extreme eigenvalues of M,
where $M_{ij} = J^-_{ij} - J^+_{ij} - \gamma\delta_{ij}$;
$V^0_{ia} = \frac{1}{K} + \varepsilon \cdot r$ *and* $[0, 1]$ $\forall i, a$;
REPEAT
•REPEAT-2
$\forall i$ Do:

$$H^s_{ia} = \sum_{j=1}^{N} (J^-_{ij} - J^+_{ij}) V_{ja}^{\begin{cases} s+1, & j<i \\ s, & j\geqslant i \end{cases}} - \gamma V^s_{ia} \quad \forall a;$$

$$F^s_i = \sum_{a=1}^{K} \mathrm{e}^{-H^s_{ia}} / T^t;$$

$$V^{s+1}_{ia} = \frac{\mathrm{e}^{-H^s_{ia}/T^t}}{F^s_i} + \varepsilon \cdot r \text{ and } [0, 1] \ \forall a;$$

$$s = s + 1;$$

UNTIL-2

$$\frac{1}{N} \sum_{i,a} |V^s_{ia} - V^{s-1}_{ia}| \leqslant 0.01;$$

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

$$\frac{1}{N} \sum_{i,a} (V^s_{ia})^2 \geqslant 0.99;$$

RETURN

$$\{\chi_a | \forall S_i \in \chi_a \cdot \forall b \neq a \ V^s_{ia} > V^s_{ib}\};$$

5.2. A quantitative experiment

Let us now study a larger quantitative experiment. While it is possible to cluster thousands of belief functions [10] with the Potts spin clustering method, we will study a smaller yet challenging problem where we easily can observe the characteristics of clustering SSFs and ISSFs and in detail study the interaction between conflicts and constraints in the clustering process.

We assume a frame of discernment with four focal elements $\Theta = \{x, y, z, q\}$ and a problem with four belief functions with uniformly drawn random mass assignments for all 15 focal elements in a way such that

$$\sum_{\emptyset \neq X \subseteq \Theta} m(X) = 1, \tag{28}$$

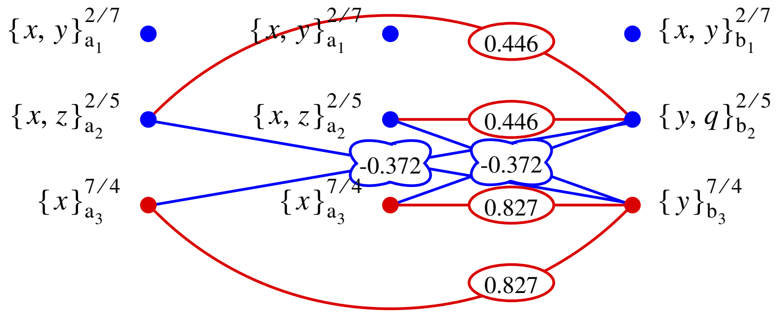

Fig. 3. Generalized weights of conflict J^-.

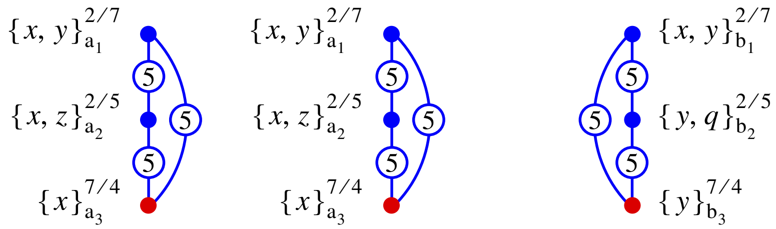

Fig. 4. Weights of attractions J^+ (constraints). Here $\beta = 5$.

with $m(\emptyset) = 0$. Each of these four belief functions is then decomposed into 16 SSFs or ISSFs, one for each subset of the frame. This is done using the decomposition algorithm available in TBMLAB [20]. Here one of the 16 SSFs or ISSFs support $m(\emptyset)$.

The 64 SSFs and ISSFs resulting from the decomposition of the four belief functions are clustered into four clusters using the Potts spin neural clustering method. In the illustration in Fig. 5 the four 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 (not shown). The vector is scaled by the absolute value of the output signal. As each SSF and ISSF have partial memberships with four clusters during the clustering process four such vectors correspond to one SSF or ISSF. At every iteration of the neural network each SSF and ISSF is plotted as a point in the normal plane of the cylinder's symmetry axis as the sum of these four vectors and by iteration step on the vertical axis. Together these points from different iterations make up a path shown as a line in Fig. 5 (starting from the center of the circle at iteration step 1 and terminating at one of the four cluster positions at the final iteration step). A point along the path is a visualization of the weighted average of the four partial cluster memberships for the SSF or ISSF during the convergence. One such path is plotted for each of the 64 SSFs or ISSFs. As seen in Fig. 5 most of the convergence take place at the tenth iteration step. This was done with a $\beta = 0.17$, the lowest β to achieve one or more perfect clusterings from 100 trials.

In Fig. 6 we see a top view of the same clustering process. Notice that most of the 64 SSFs and ISSFs head directly for an appropriate cluster, only a handful have to change course from one cluster to another as the process converges.

In Fig. 7 we observe the convergence of the process. The curve in the figure is an entropy-like measure

$$-\sum_{ia} V_{ia} \ln(V_{ia}) \tag{29}$$

summed up over all output signals from the 256 neurons, corresponding to the partial memberships of the 64 SSFs and ISSFs in each of the four clusters at each iteration step. The entropy starts at 88.723 $(64 \times 4 \times [-0.25 \ln 0.25])$ and drops quickly from the ninth to the tenth iteration followed by a final steady convergence.

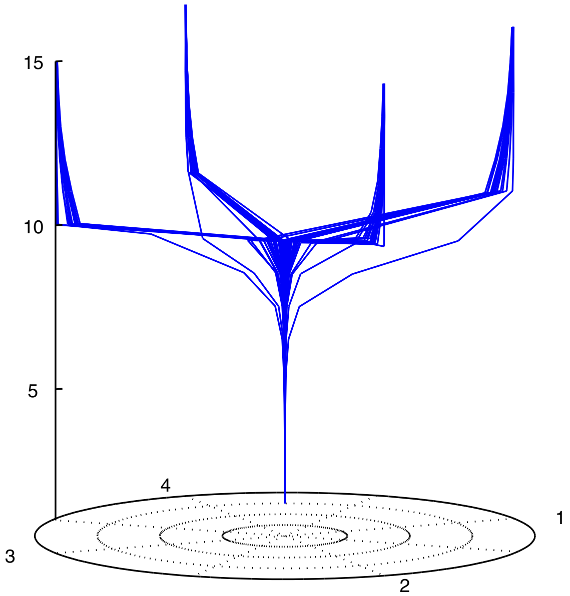

Fig. 5. The clustering process of 64 SSFs and ISSFs into four clusters using both conflicts and constraints as attractions. Here $\beta = 0.17$ and convergence is achieved in 15 iterations. There are no clustering errors in this example, i.e., all SSFs and ISSFs decomposed from one belief function also end up in the same cluster, and no SSFs or ISSFs from any of the other three belief functions are ever in that cluster.

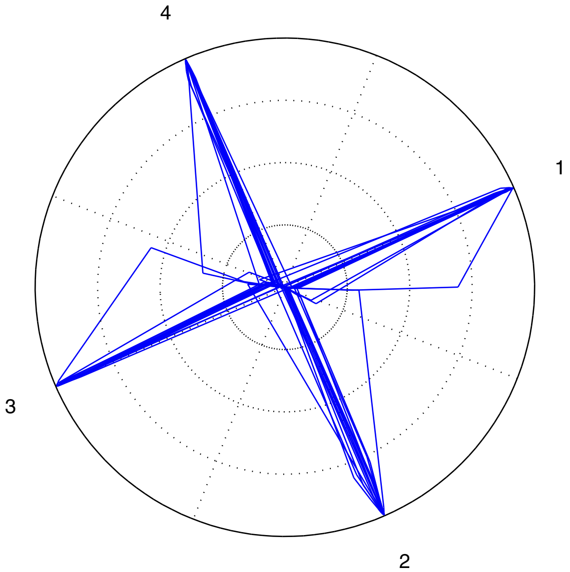

Fig. 6. A top view of Fig. 5. Most of the 64 SSFs and ISSFs head straight from the center towards their corresponding cluster.

In Fig. 8 we measure the average clustering error over 100 runs as a function of β. While the earlier observations and figures were aimed at understanding the behavior of the clustering process, this is a measure of overall performance of the process. The average clustering error is calculated as the number of SSFs and ISSFs that are misplaced into another cluster relative to its corresponding belief function. The cluster corresponding to the belief function in question is determined as the cluster with the most SSFs and ISSFs from this belief function. This is always determined in this indirect way as no clusters are pre-labeled before the clustering has taken place. When all 16 SSFs and ISSFs that were decomposed from one belief function end up in the

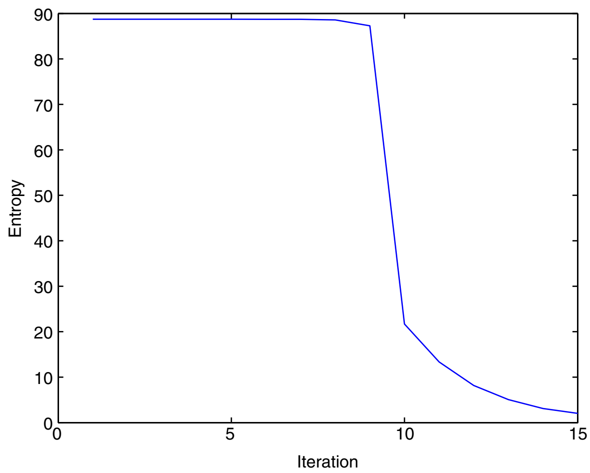

Fig. 7. The convergence of the clustering process over 15 iterations. The curve is the sum of an entropy-like measure of 256 neurons (64×4) corresponding to the 64 SSFs and ISSFs of the scattering of their partial membership to the four clusters.

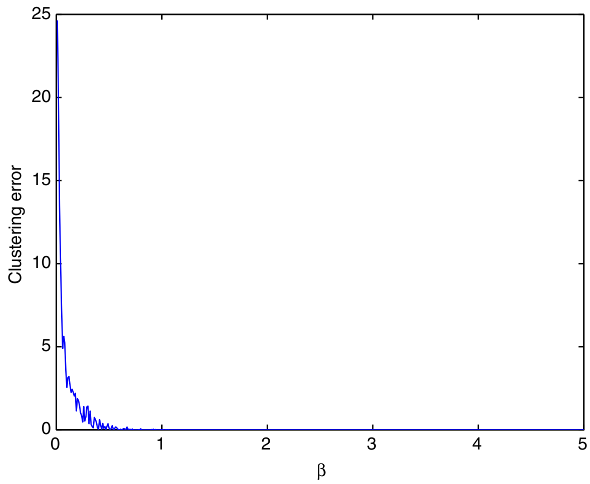

Fig. 8. Clustering error as a function of β (average over 100 runs).

same cluster, one gets a clustering error of zero for that belief function. When this is achieved for all four belief functions, one gets an overall clustering error of zero. The maximum clustering error is when the 16 SSFs and ISSFs are evenly distributed among the four clusters. Any cluster may now be said to correspond to the belief function for a clustering error of 12 (3×4) for this belief function, and a maximum overall clustering error of 48 (4×12).

We observe in Fig. 8 that the average clustering error over 100 clustering processes for each value of β quickly drops from 24.64 for $\beta = 0.01$ towards zero for β values around 1. The lowest β with clustering errors of zero for all 100 runs is 0.62. For $\beta \geqslant 1.07$ there are never any clustering errors.

In Fig. 9 we observe the average clustering computation time over 100 runs for all β values from 0.01 until 5.00. The highest computation time was 0.1039 s for $\beta = 0.01$. At $\beta = 1$ we obtained a computation time of 0.0456 s. The computation time is quite stable for larger β values.

For larger problem sizes we have shown [2] that the Potts spin clustering computation time scales as n^2k^2, where n are the number of SSFs and ISSFs and k the number of clusters. If the number of belief functions are k with $n = 2^k$ SSFs and ISSFs, this yields a computational complexity of $O(2^{2k}k^2)$ when measured by the num-

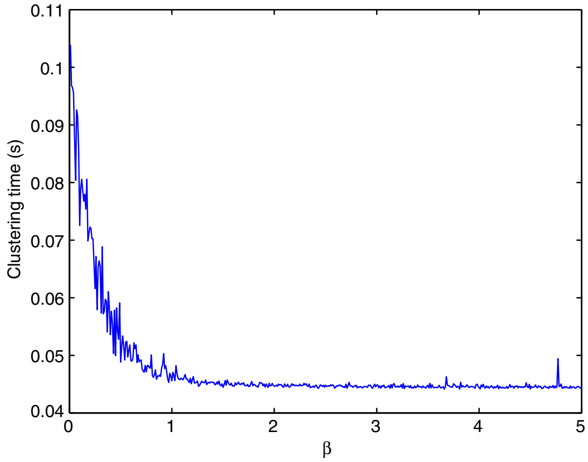

Fig. 9. Clustering time as a function of β (average over 100 runs). All times are measured in seconds, running MATLAB Version 7.3 (R2006b) on an Intel Xeon processor (2.80 GHz CPU, 2 GB RAM).

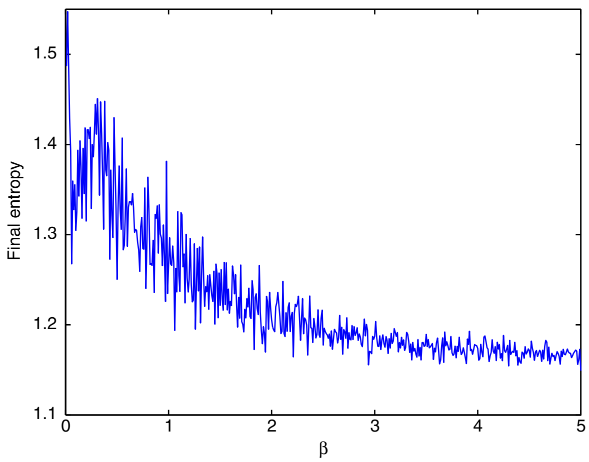

Fig. 10. Final entropy after convergence as a function of β (average over 100 runs).

ber of belief functions or $O(n^2\log^2 n)$ when measured by the number of SSFs and ISSFs, i.e., in the number of inputs to the clustering process.

Finally, in Fig. 10 we observe the final remaining entropy after convergence is achieved. We notice a slightly better convergence with less remaining scattering for higher β values above three, although the results are fairly good for any $\beta \geqslant 1$.

Thus, from both a clustering performance and a computational point of view we should prefer β values larger than 1. Although of lesser importance, we also notice the slightly better convergence of the clustering process itself when $\beta \geqslant 3$.

6. Conclusions

In this paper, we have developed a methodology which makes it possible to cluster belief functions that are mixed up. The belief functions are first decomposed into simple support functions and inverse simple support functions. We then adopt a neural clustering algorithm intended for simple support functions to handle both

SSFs and ISSFs while recording their decomposition for postclustering recomposing. With this method we may cluster any type of belief function, and in particular non-consonant belief functions.

Acknowledgements

I wish to express my sincere appreciation to the late Prof. Philippe Smets. The idea to decompose mixed up belief functions into SSFs and ISSFs in order to cluster all pieces was suggested by him [21].

References

[1] J. Schubert, Managing decomposed belief functions, in: Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, Paris, France, 2006, vol. 2, pp. 1428–1435.
[2] M. Bengtsson, J. Schubert, Dempster–Shafer clustering using Potts spin mean field theory, Soft Computing 5 (3) (2001) 215–228.
[3] J. Schubert, On nonspecific evidence, International Journal of Intelligent Systems 8 (6) (1993) 711–725.
[4] J. Schubert, Specifying nonspecific evidence, International Journal of Intelligent Systems 11 (8) (1996) 525–563.
[5] J. Schubert, Managing inconsistent intelligence, in: Proceedings of the 3rd International Conference on Information Fusion, Paris, France, 2000, vol. 1, pp. TuB4/10–16.
[6] A.P. Dempster, A generalization of Bayesian inference, Journal of the Royal Statistical Society Series B 30 (2) (1968) 205–247.
[7] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[8] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (2) (1994) 191–234.
[9] 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.
[10] J. Schubert, Clustering belief functions based on attracting and conflicting metalevel evidence using Potts spin mean field theory, Information Fusion 5 (4) (2004) 309–318.
[11] P. Smets, The canonical decomposition of a weighted belief, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montréal, Canada, 1995, vol. 2, pp. 1896–1901.
[12] J. Schubert, H. Sidenbladh, Sequential clustering with particle filtering estimating the number of clusters from data, in: Proceedings of the 8th International Conference on Information Fusion, Philadelphia, PA, 2005, Paper A4-3, pp. 1–8.
[13] S. Salvador, P. Chan, Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms, in: Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence, Boca Raton, FL, 2004, pp. 576–584.
[14] Informationssystem Flygvapnet version 1.0 (IS FV II 1.0A), Swedish Armed Forces, 2002.
[15] S. Ahlberg, P. Hörling, K. Johansson, K. Jöred, H. Kjellström, C. Mårtenson, G. Neider, J. Schubert, P. Svenson, P. Svensson, J. Walter, An information fusion demonstrator for tactical intelligence processing in network-based defense, Information Fusion 8 (1) (2007) 84–107.
[16] P. Smets, Analyzing the combination of conflicting belief functions, Information Fusion, in press, doi:10.1016/j.inffus.2006.04.003.
[17] R. Kennes, Computational aspects of the Möbius transformation of graphs, IEEE Transactions on Systems, Man, and Cybernetics 22 (2) (1992) 201–223.
[18] F.Y. Wu, The Potts model, Reviews of Modern Physics 54 (1) (1982) 235–268.
[19] C. Peterson, B. Söderberg, A new method for mapping optimization problems onto neural networks, International Journal of Neural Systems 1 (1) (1989) 3–22.
[20] P. Smets, TBMLAB 3.1, 2004, March. [Online]. Available: http://iridia.ulb.ac.be/~psmets/#G.
[21] P. Smets, Personal communication, 2000.