December 2008 Volume 3 Number 2 ISSN 1557-6418

JAIF 08

Journal of Advances in Information Fusion

A semi-annual archival publication of the International Society of Information Fusion

Regular Papers Page

ISIF

INTERNATIONAL SOCIETY OF INFORMATION FUSION

The International Society of Information Fusion (ISIF) seeks to be the premier professional society and global information resource for multidisciplinary approaches for theoretical and applied INFORMATION FUSION technologies. Technical areas of interest include target tracking, detection theory, applications for information fusion methods, image fusion, fusion systems architectures and management issues, classification, learning, data mining, Bayesian and reasoning methods.

Manuscripts are submitted at http://jaif.msubmit.net. If in doubt about the proper editorial area of a contribution, submit it under the unknown area.

Journal of Information Advances in Information Fusion (ISSN 1557-6418) is published semi-annually by the International Society of Information Fusion. The responsibility for the contents rests upon the authors and not upon ISIF, the Society, or its members. ISIF is California Nonprofit Public Benefit Corporation at P.O. Box 4631, Mountain View, California 94040.

From the Associate Editor-in-Chief

December 2008

Makings of a Good Peer Review

By accepting to be a reviewer, one takes on the responsibility to read the paper thoroughly and provide detailed comments and explanations for rejecting, revising, or accepting a paper. It is also important for a reviewer to maintain scientific objectivity so that the decision or score for a paper is justified. Some of the problems that I have observed in our reviews are very similar to those experienced by the famous physicist late Richard Feynman while he was reviewing new school books for California.

While at Caltech, Feynman worked in the California State Curriculum Commission for the Los Angeles area (excluding the city of Los Angeles) that chose new school books for California. The book depository sent him a large number of books weighing about three hundred pounds. The book review process was a daunting task. However, he read all the books carefully and kept detailed notes of the review. He worked hard in the basement of his house to read the books. His wife would say that during this review period, "it was like living over a volcano. It would be quiet for a while, but then all of a sudden, *Blllllooooowwwww!!!!*—there would be a big explosion from the *volcano* below [1]." The reason for Feynman's reaction was that the books were very poorly written.

Feynman went to his first meeting with other members of the commission. Some members of the commission asked him what he thought about a certain book. He replied that he didn't receive the book from the depository (because the book was not complete) and therefore, he didn't have a judgment on it [1]. However, some other members of the commission had a rating on this missing book and in fact; the rating on the missing book was a little bit higher than two other books in a set of three books [1].

Kuo-Chu Chang of George Mason University and I served as Technical Co-chairs for the Fusion 2009 conference. From the peer review process for Fusion 2009 papers, I observed that some reviewers did not read the

papers thoroughly and gave high scores without sufficient explanations. This tends to happen when the session chair of a special session reviews papers from authors that he/she has contacted for the special session. Reviewers associated with the topic of the special session also give a high score to a paper when it doesn't deserve that score. This causes a number of problems. Some weak papers may get high scores and be considered for an award. Secondly, papers near the borderline between acceptance and rejection can be rejected. Therefore, it is extremely important for the reviewers to be scientifically objective and for the Technical Chairs to assign the papers appropriately such that these problems can be avoided. We have also observed that some reviewers have not spent a sufficient amount of time and have done a hasty review.

Bieber [2] provides a number of useful and practical suggestions for reviewing a conference or journal paper. He recommends reading the submission three times: "the first to get a feel for it, the second reading the paper in depth, and the third to actually mark it up." He suggests filling out the refereeing form right after the third reading, while things are still fresh in memory.

Halmos [3] stresses "Honesty is the Best Policy" in mathematical writing. The same principle also applies while reviewing a paper. If a reviewer gives a high score to a paper without reading the paper thoroughly, then it violates this principle. If a reviewer does not have time to review a paper carefully, then it is best for the reviewer to notify the editors about it. Thorough and fair reviewing does not go unnoticed by editors, and it can establish a good reputation for a reviewer, which can also spread in the research community [2].

At times, I have noticed that when an associate editor of the JAIF recommends a paper for publication and it arrives for final proof reading, obvious errors related to figures, tables, equations, and references are found. It is quite surprising that such errors have gone unnoticed by all of the reviewers. Technical errors as well as these errors should have been identified by the reviewers and addressed by the authors and checked by the associate editor before it was recommended for publication.

The success of JAIF strongly depends on high quality, fair, and timely reviews. Therefore, I encourage ISIF members and non-members to contact the Editor-In-Chief or me to be reviewers for JAIF.

Mahendra Mallick
Associate Editor-in-chief

[1] Feynman, R. P.
"Surely You're Joking Mr. Feynman!": Adventures of a Curious Character, New York, W. W. Norton & Company, 1985.

[2] Bieber, M.
"How to Review," http://web.njit.edu/∼bieber/review.html.

[3] Halmos, P. R.
In *How to Write Mathematics*, American Mathematical Society, 1973, 19–48.

General Combination Rules for Qualitative and Quantitative Beliefs

ARNAUD MARTIN
CHRISTOPHE OSSWALD
JEAN DEZERT
FLORENTIN SMARANDACHE

Martin and Osswald [15] have recently proposed many generalizations of combination rules on quantitative beliefs in order to manage the conflict and to consider the specificity of the responses of the experts. Since the experts express themselves usually in natural language with linguistic labels, Smarandache and Dezert [13] have introduced a mathematical framework for dealing directly also with qualitative beliefs. In this paper we recall some element of our previous works and propose the new combination rules, developed for the fusion of both qualitative or quantitative beliefs.

Manuscript received September 24, 2007; released for publication August 5, 2008.

Refereeing of this contribution was handled by Fabio Roli.

Authors' addresses: Arnaud Martin and Christophe Osswald, E^3I^2 EA3876, ENSIETA, 2 rue François Verny, 29806 Brest Cedex 09, France, E-mail: (Arnaud.Martin,Christophe.Osswald@ensieta.fr); Jean Dezert, ONERA, The French Aerospace Lab, 29 Av. Division Leclerc, 92320 Châtillon, France, E-mail: (jean.dezert@onera.fr); Florentin Smarandache, Chair of Math. & Sciences Dept., University of New Mexico, 200 College Road, Gallup, NM 87301, U.S.A., E-mail: (smarand@unm.edu).

1557-6418/08/$17.00

1. INTRODUCTION

Many fusion theories have been studied for the combination of the experts opinions expressed either quantitatively or qualitatively such as voting rules [11], [31], possibility theory [6], [35], and belief functions theory [2], [17]. All these fusion approaches can be divided basically into four steps: *modeling*, *parameters estimation* (depending on the model, not always necessary)), *combination* and *decision*. The most difficult step is presumably the first one which depends highly on the problem and application we have to cope with. However, it is only at the combination step that we can take into account useful information such as the conflict (partial or total) between the experts and/or the specificity of the expert's response.

The voting rules are not adapted to the modeling of conflict between experts [31]. Although both possibility and probability-based theories can model imprecise and uncertain data at the same time, in many applications, the experts are only able to express their "certainty" (or belief) only from their partial knowledge, experience and from their own perception of the reality. In such context, the belief function-based theories provide an appealing general mathematical framework for dealing with quantitative and qualitative beliefs.

In this paper we present the most recent advances in belief functions theory for managing the conflict between the sources of evidence/experts and their specificity. For the first time in the literature both the quantitative and qualitative aspects of beliefs are presented in a unified mathematical framework. This paper actually extends the work in two papers [13], [15] presented during the 10th International Conference on Information Fusion (Fusion 2007) in Québec City, Canada on July 9–12, 2007 in the session "Combination in Evidence Theory."

Section 2 briefly recalls the basis of belief functions theories, *i.e.* the Mathematical Theory of Evidence or Dempster-Shafer theory (DST) developed by Shafer in 1976 [2], [17], and its natural extension called Dezert-Smarandache Theory (DSmT) [18], [19], [20]. We introduce in this section the notion of quantitative and qualitative beliefs and the operators on linguistic labels for dealing directly with qualitative beliefs. Section 3 presents the main classical quantitative combination rules used so far, *i.e.* Dempster's rule, Yager's rule, Dubois-Prade's rule and the recent Proportional Conflict Redistribution rules (PCR) proposed by Smarandache and Dezert [22] and extended by Martin and Osswald in [19]. Some examples are given to illustrate how these rules work. Section 4 explains through different examples how all the classical quantitative combination rules can be directly and simply translated/extended into the qualitative domain in order to combine easily any qualitative beliefs expressed in natural language by linguistic labels. Section 5 proposes new general quantitative rules of combination which allow to take into

account both the discounting of the sources (if any) and the proportional conflict redistribution. The direct extension of these general rules into the qualitative domain is then presented in details on several examples in Section 6.

2. BASIS OF DST AND DSmT

A. Power Set and Hyper-Power Set

In DST framework, one considers a frame of discernment $\Theta = \{\theta_1, \ldots, \theta_n\}$ as a finite set of n exclusive and exhaustive elements (*i.e.* Shafer's model denoted $\mathcal{M}^0(\Theta)$). The *power set* of Θ is the set of all subsets of Θ. The order of a power set of a set of order/cardinality $|\Theta| = n$ is 2^n. The power set of Θ is denoted 2^Θ. For example, if $\Theta = \{\theta_1, \theta_2\}$, then $2^\Theta = \{\emptyset, \theta_1, \theta_2, \theta_1 \cup \theta_2\}$.

In DSmT framework, one considers $\Theta = \{\theta_1, \ldots, \theta_n\}$ be a finite set of n exhaustive elements only (*i.e.* free DSm-model denoted $\mathcal{M}^f(\Theta)$). Eventually some integrity constraints can be introduced in this free model depending on the nature of the problem of interest. The *hyper-power set* of Θ (*i.e.* the free Dedekind's lattice) denoted D^Θ [18] is defined as

1) $\emptyset, \theta_1, \ldots, \theta_n \in D^\Theta$,
2) If $A, B \in D^\Theta$, then $A \cap B$, $A \cup B \in D^\Theta$,
3) No other elements belong to D^Θ, except those obtained by using rules 1 or 2.

If $|\Theta| = n$, then $|D^\Theta| \leq 2^{2^n}$. Since for any finite set Θ, $|D^\Theta| \geq |2^\Theta|$, we call D^Θ the *hyper-power set* of Θ. For example, if $\Theta = \{\theta_1, \theta_2\}$, then $D^\Theta = \{\emptyset, \theta_1 \cap \theta_2, \theta_1, \theta_2, \theta_1 \cup \theta_2\}$. The *free DSm model* $\mathcal{M}^f(\Theta)$ corresponding to D^Θ allows to work with vague concepts which exhibit a continuous and relative intrinsic nature. Such concepts cannot be precisely refined in an absolute interpretation because of the unreachable universal truth.

It is clear that Shafer's model $\mathcal{M}^0(\Theta)$ which assumes that all elements of Θ are truly exclusive is a more constrained model than the free-DSm model $\mathcal{M}^f(\Theta)$ and the power set 2^Θ can be obtained from hyper-power set D^Θ by introducing in $\mathcal{M}^f(\Theta)$ all exclusivity constraints between elements of Θ. Between the free-DSm model $\mathcal{M}^f(\Theta)$ and Shafer's model $\mathcal{M}^0(\Theta)$, there exists a wide class of fusion problems represented in term of the DSm hybrid models denoted $\mathcal{M}(\Theta)$ where Θ involves both fuzzy continuous and discrete hypotheses. The main differences between DST and DSmT frameworks are (i) the model on which one works with, and (ii) the choice of the combination rule and conditioning rules [18], [19]. In the sequel, we use the generic notation G^Θ for denoting either D^Θ (when working in DSmT with free DSm model) or 2^Θ (when working in DST with Shafer's model).

B. Quantitative Basic Belief Assignment (BBA)

The (quantitative) basic belief assignment (BBA) $m(\cdot)$ has been introduced for the first time in 1976 by Shafer [17] in his Mathematical Theory of Evidence (*i.e.* DST). $m(\cdot)$ is defined as a mapping function from $2^\Theta \to [0,1]$ provided by a given source of evidence $\mathcal{B}$ satisfying the conditions

$$m(\emptyset) = 0, \tag{1}$$

$$\sum_{A \in 2^\Theta} m(A) = 1. \tag{2}$$

The elements of 2^Θ having a strictly positive mass are called *focal elements* of $\mathcal{B}$. The set of focal elements of $m(\cdot)$ is called the core of $m(\cdot)$ and is usually denoted $\mathcal{F}(m)$. The equation (1) corresponds to the closed-world assumption [17]. As introduced by Smets [25], we can also define the belief function only with

$$\sum_{A \in 2^\Theta} m(A) = 1 \tag{3}$$

and thus we can have $m(\emptyset) > 0$, working with the open-world assumption. In order to change an open world to a closed world, we can always add one extra closure element in the open discriminant space Θ. In the following, we assume that we always work within a closed-world Θ.

The (quantitative) basic belief assignment (BBA) $m(\cdot)$ can also be defined similarly in the DSmT framework by working on hyper-power set D^Θ instead on classical power-set 2^Θ as within DST. More generally for taking into account some integrity constraints on (closed-world) Θ (if any), $m(\cdot)$ can be defined on G^Θ as

$$m(\emptyset) = 0, \tag{4}$$

$$\sum_{A \in G^\Theta} m(A) = 1. \tag{5}$$

The conditions (1)–(5) give a large panel of definitions of the belief functions, which is one of the difficulties of the theories. From any basic belief assignments $m(\cdot)$, other belief functions can be defined such as the credibility Bel$(\cdot)$ and the plausibility Pl$(\cdot)$ [17], [18] which are in one-to-one correspondence with $m(\cdot)$.

After combining several BBAs provided by several sources of evidence into a single one with some chosen fusion rule (see next section), one usually has also to make a final decision to select the "best" hypothesis representing the unknown truth for the problem under consideration. Several approaches are generally adopted for decision-making from belief functions $m(\cdot)$, Bel$(\cdot)$ or Pl$(\cdot)$. The maximum of the credibility function Bel$(\cdot)$ is known to provide a pessimistic decision, while the maximum of the plausibility function Pl$(\cdot)$ is often considered as too optimistic. A common solution for decision-making in these frameworks is to use the *pignistic probability* denoted BetP(X) [25] which offers a good compromise between the max of Bel$(\cdot)$ and the max of Pl$(\cdot)$. The pignistic probability in DST frame-

work is given for all $X \in 2^\Theta$, with $X \neq \emptyset$ by

$$\mathrm{BetP}(X) = \sum_{Y \in 2^\Theta, Y \neq \emptyset} \frac{|X \cap Y|}{|Y|} \frac{m(Y)}{1 - m(\emptyset)}, \tag{6}$$

for $m(\emptyset) \neq 1$. The pignistic probability can also be defined in DSmT framework as well (see Chapter 7 of [18] for details).

When we can quantify/estimate the reliability of each source of evidence, we can weaken the basic belief assignment before the combination by the classical discounting procedure [17]

$$\begin{aligned} m'(X) &= \alpha m(X), \qquad \forall \quad X \in 2^\Theta \setminus \{\Theta\} \\ m'(\Theta) &= \alpha m(\Theta) + 1 - \alpha, \end{aligned} \tag{7}$$

where $\alpha \in [0,1]$ is the discounting factor of the source of evidence $\mathcal{B}$ that is in this case the reliability of the source of evidence $\mathcal{B}$, eventually as a function of $X \in 2^\Theta$. Same procedure can be applied for BBAs defined on G^Θ in DSmT framework.

C. Qualitative Basic Belief Assignment (QBBA)

1) *Qualitative operators on linguistic lables*

Recently Smarandache and Dezert [13], [19] have proposed an extension of classical quantitative belief assignments and numerical operators to qualitative beliefs expressed by linguistic labels and qualitative operators in order to be closer to what human experts can easily provide. In order to compute directly with words/linguistic labels and qualitative belief assignments instead of quantitative belief assignments over G^Θ, Smarandache and Dezert have defined in [19] a *qualitative basic belief assignment* $qm(\cdot)$ as a mapping function from G^Θ into a set of linguistic labels $L = \{L_0, \tilde{L}, L_{n+1}\}$ where $\tilde{L} = \{L_1, \ldots, L_n\}$ is a finite set of linguistic labels and where $n \geq 2$ is an integer. For example, L_1 can take the linguistic value "poor," L_2 the linguistic value "good," etc. $\tilde{L}$ is endowed with a total order relationship $\prec$, so that $L_1 \prec L_2 \prec \cdots \prec L_n$. To work on a true closed linguistic set L under linguistic addition and multiplication operators, Smarandache and Dezert extended naturally $\tilde{L}$ with two extreme values $L_0 = L_{\min}$ and $L_{n+1} = L_{\max}$, where L_0 corresponds to the minimal qualitative value and L_{n+1} corresponds to the maximal qualitative value, in such a way that $L_0 \prec L_1 \prec L_2 \prec \cdots \prec L_n \prec L_{n+1}$, where $\prec$ means inferior to, less (in quality) than, or smaller than, etc. Labels $L_0, L_1, L_2, \ldots, L_n, L_{n+1}$ are called *linguistically equidistant* if: $L_{i+1} - L_i = L_i - L_{i-1}$ for all $i = 1, 2, \ldots, n$ where the definition of subtraction of labels is given in the sequel by (14). In the sequel $L_i \in L$ are assumed linguistically equidistant[1] labels such that we can make an isomorphism between $L = \{L_0, L_1, L_2, \ldots, L_n, L_{n+1}\}$ and $\{0, 1/(n+1), 2/(n+1), \ldots, n/(n+1), 1\}$, defined as $L_i = i/(n+1)$ for all $i = 0, 1, 2, \ldots, n, n+1$. Using this isomorphism, and making an analogy to the classical operations of real numbers, we are able to justify and define precisely the following qualitative operators (or q-operators for short).

- q-addition of linguistic labels

$$L_i + L_j = \frac{i}{n+1} + \frac{j}{n+1} = \frac{i+j}{n+1} = L_{i+j}, \tag{8}$$

  we set the restriction that $i + j < n + 1$; in the case when $i + j \geq n + 1$ we restrict $L_{i+j} = L_{n+1} = L_{\max}$. This is the justification of the qualitative addition we have defined.
- q-multiplication of linguistic labels[2]
  a) Since

$$L_i \cdot L_j = \frac{i}{n+1} \cdot \frac{j}{n+1} = \frac{(i \cdot j)/(n+1)}{n+1},$$

  the best approximation would be $L_{[(i \cdot j)/(n+1)]}$, where $[x]$ means the closest integer to x (with $[n + 0.5] = n + 1$, $\forall n \in \mathbb{N}$), *i.e.*

$$L_i \cdot L_j = L_{[(i \cdot j)/(n+1)]}. \tag{9}$$

  For example, if we have L_0, L_1, L_2, L_3, L_4, L_5, corresponding to respectively 0, 0.2, 0.4, 0.6, 0.8, 1, then $L_2 \cdot L_3 = L_{[(2 \cdot 3)/5]} = L_{[6/5]} = L_{[1.2]} = L_1$; using numbers: $0.4 \cdot 0.6 = 0.24 \approx 0.2 = L_1$; also $L_3 \cdot L_3 = L_{[(3 \cdot 3)/5]} = L_{[9/5]} = L_{[1.8]} = L_2$; using numbers $0.6 \cdot 0.6 = 0.36 \approx 0.4 = L_2$.
  b) A simpler approximation of the multiplication, but less accurate (as proposed in [19]) is thus

$$L_i \cdot L_j = L_{\min\{i,j\}}. \tag{10}$$

- Scalar multiplication of a linguistic label
  Let a be a real number. We define the multiplication of a linguistic label by a scalar as follows

$$a \cdot L_i = \frac{a \cdot i}{n+1} \approx \begin{cases} L_{[a \cdot i]} & \text{if} \quad [a \cdot i] \geq 0, \\ L_{-[a \cdot i]} & \text{otherwise.} \end{cases} \tag{11}$$

- Division of linguistic labels
  a) Division as an internal operator: $/ : L \cdot L \to L$. Let $j \neq 0$, then

$$L_i / L_j = \begin{cases} L_{[(i/j) \cdot (n+1)]} & \text{if} \quad [(i/j) \cdot (n+1)] < n+1, \\ L_{n+1} & \text{otherwise.} \end{cases} \tag{12}$$

  The first equality in (12) is well justified because when $[(i/j) \cdot (n+1)] < n+1$, one has

$$L_i / L_j = \frac{i/(n+1)}{j/(n+1)} = \frac{(i/j) \cdot (n+1)}{n+1} = L_{[(i/j) \cdot (n+1)]}.$$

[1] If the labels are not equidistant, the q-operators still work, but they are less accurate.

[2] The q-multiplication of two linguistic labels defined here can be extended directly to the multiplication of $n > 2$ linguistic labels. For example the product of three linguistic label will be defined as $L_i \cdot L_j \cdot L_k = L_{[(i \cdot j \cdot k)/(n+1)(n+1)]}$, etc.

For example, if we have L_0, L_1, L_2, L_3, L_4, L_5, corresponding to respectively 0, 0.2, 0.4, 0.6, 0.8, 1, then: $L_1/L_3 = L_{[(1/3)\cdot 5]} = L_{[5/3]} = L_{[1.66]} \approx L_2 . L_4/L_2 = L_{[(4/2)\cdot 5]} = L_{[2\cdot 5]} = L_{\max} = L_5$ since $10 > 5$.
b) Division as an external operator: $\oslash : L \cdot L \to \mathbb{R}^+$. Let $j \neq 0$. Since $L_i \oslash L_j = (i/(n+1))/(j/(n+1)) = i/j$, we simply define

$$L_i \oslash L_j = i/j. \tag{13}$$

Justification of b): When we divide say L_4/L_1 in the above example, we get $0.8/0.2 = 4$, but no label is corresponding to number 4 which is not included in the interval $[0,1]$, hence the division as an internal operator we need to get as a response label, so in our example we approximate it to $L_{\max} = L_5$, which is a very rough approximation! Therefore, depending on the fusion combination rules, it may be better to consider the qualitative division as an external operator, which gives us the exact result.

- q-subtraction of linguistic labels given by $- : L \cdot L \to \{L, -L\}$,

$$L_i - L_j = \begin{cases} L_{i-j} & \text{if} \quad i \geq j, \\ -L_{j-i} & \text{if} \quad i < j. \end{cases} \tag{14}$$

where $-L = \{-L_1, -L_2, \ldots, -L_n, -L_{n+1}\}$. The q-subtraction above is well justified since when $i \geq j$, one has $L_i - L_j = i/(n+1) - j/(n+1) = (i-j)/(n+1)$.

The previous qualitative operators are logical due to the isomorphism between the set of linguistic equidistant labels and a set of equidistant numbers in the interval $[0,1]$. These qualitative operators are built exactly on the track of their corresponding numerical operators, so they are more mathematically defined than the ad-hoc definitions of qualitative operators proposed in the literature so far. The extension of these operators for handling quantitative or qualitative enriched linguistic labels can be found in [13].

Remark about doing multi-operations on labels

When working with labels, no matter how many operations we have, the best (most accurate) result is obtained if we do only one approximation. That one should be at the end. For example, if we have to compute terms like $L_i L_j L_k/(L_p + L_q)$ as for qualitative proportional conflict redistribution (QPCR) rule (see example in Section 4), we compute all operations as defined above. Without any approximations (*i.e.* not even calculating the integer part of indexes, neither replacing by $n+1$ if the intermediate results is bigger than $n+1$). Then

$$\begin{aligned} \frac{L_i L_j L_k}{L_p + L_q} &= \frac{L_{(ijk)/(n+1)^2}}{L_{p+q}} \\ &= L_{\frac{(ijk)/(n+1)^2}{p+q}\cdot(n+1)} \\ &= L_{\frac{(ijk)/(n+1)}{p+q}} = L_{\frac{ijk}{(n+1)(p+q)}}, \end{aligned} \tag{15}$$

and now, when all work is done, we compute the integer part of the index, *i.e.* $[ijk/((n+1)(p+q))]$ or replace it by $n+1$ if the final result is bigger than $n+1$. Therefore, the term $L_i L_j L_k/(L_p + L_q)$ will take the linguistic value L_{n+1} whenever $[ijk/((n+1)(p+q))] > n+1$. This method also insures us of a unique result, and it is mathematically closer to the result that would be obtained if working with corresponding numerical masses. Otherwise, if one approximates either at the beginning or end of each operation or in the middle of calculations, the inaccuracy propagates (becomes bigger) and we obtain different results, depending on the places where the approximations were done. If we need to round the labels' indexes to integer indexes, for a better accuracy of the result, this rounding must be done at the very end. If we work with fractional/decimal indexes (therefore no approximations), then we can normally apply the qualitative operators one by one in the order they are needed; in this way the quasi-normalization is always kept.

2) *Quasi-normalization of $qm(\cdot)$*

There is no known way to define a normalized $qm(\cdot)$, but a qualitative quasi-normalization [19], [24] is nevertheless possible when considering equidistant linguistic labels because in such case, $qm(X_i) = L_i$, is equivalent to a quantitative mass $m(X_i) = i/(n+1)$ which is normalized if

$$\sum_{X \in G^\Theta} m(X) = \sum_k i_k/(n+1) = 1,$$

but this one is equivalent to

$$\sum_{X \in G^\Theta} qm(X) = \sum_k L_{i_k} = L_{n+1}.$$

In this case, we have a *qualitative normalization*, similar to the (classical) numerical normalization. However, if the previous labels $L_0, L_1, L_2, \ldots, L_n, L_{n+1}$ from the set L are not equidistant, the interval $[0,1]$ cannot be split into equal parts according to the distribution of the labels. Then it makes sense to consider a *qualitative quasi-normalization*, *i.e.* an approximation of the (classical) numerical normalization for the qualitative masses in the same way

$$\sum_{X \in G^\Theta} qm(X) = L_{n+1}.$$

In general, if we don't know if the labels are equidistant or not, we say that a qualitative mass is quasi-normalized when the above summation holds. In the sequel, for simplicity, one assumes to work with quasi-normalized qualitative basic belief assignments.

From these very simple qualitative operators, it is possible to extend directly all the quantitative combination rules to their qualitative counterparts as we will show in the sequel.

3) *Working with refined labels*

- We can further extend the standard labels (those with positive integer indexes) to refined labels, *i.e.* labels with fractional/decimal indexes. In such a way, we get a more exact result, and the quasi-normalization is kept.
  Consider a simple example: If $L_2 = good$ and $L_3 = best$, then $L_{2.5} = better$, which is a qualitative (a refined label) in between L_2 and L_3.
- Further, we consider the confidence degree in a label, and give more interpretations/approximations to the qualitative information.
  For example: $L_{2/5} = (1/5) \cdot L_2$, which means that we are 20% confident in label L_2; or $L_{2/5} = (2/5) \cdot L_1$, which means that we are 40% confident in label L_1, so L_1 is closer to reality than L_2; we get 100% confidence in $L_{2/5} = 1 \cdot L_{2/5}$.

4) *Working with non-equidistant labels*: We are not able to find (for non-equidistant labels) exact corresponding numerical values in the interval $[0,1]$ in order to reduce the qualitative fusion to a quantitative fusion, but only approximations. We, herfore, prefer the use of labels.

3. CLASSICAL QUANTITATIVE COMBINATION RULES

The normalized conjunctive combination rule also called Dempster-Shafer (DS) rule is the first rule proposed in the belief theory by Shafer following Dempster's works in sixties [2]. In the belief functions theory one of the major problems is the conflict repartition enlightened by the famous Zadeh's example [36]. Since Zadeh's paper, many combination rules have been proposed, building a solution to this problem [4], [5], [7]–[9], [14], [21], [26], [27], [34]. In recent years, some unification rules have been proposed [1], [12], [29]. We briefly browse the major rules developed and used in the fusion community working with belief functions through last past thirty years (see [30] and [19] for a more comprehensive survey).

To simplify the notations, we consider only two independent sources of evidence $\mathcal{B}_1$ and $\mathcal{B}_2$ over the same frame Θ with their corresponding BBAs $m_1(\cdot)$ and $m_2(\cdot)$. Most of the fusion operators proposed in the literature use either the conjunctive operator, the disjunctive operator or a particular combination of them. These operators are respectively defined $\forall A \in G^\Theta$, by

$$m_\vee(A) = (m_1 \vee m_2)(A) = \sum_{\substack{X,Y \in G^\Theta \\ X \cup Y = A}} m_1(X) m_2(Y), \tag{16}$$

$$m_\wedge(A) = (m_1 \wedge m_2)(A) = \sum_{\substack{X,Y \in G^\Theta \\ X \cap Y = A}} m_1(X) m_2(Y). \tag{17}$$

The global/total *degree of conflict* between the sources $\mathcal{B}_1$ and $\mathcal{B}_2$ is defined by

$$k \triangleq m_\wedge(\emptyset) = \sum_{\substack{X,Y \in G^\Theta \\ X \cap Y = \emptyset}} m_1(X) m_2(Y). \tag{18}$$

If k is close to 0, the BBAs $m_1(\cdot)$ and $m_2(\cdot)$ are almost not in conflict, while if k is close to 1, the BBAs are almost in total conflict. Next, we briefly review the main common quantitative fusion rules encountered in the literature and used in engineering applications.

EXAMPLE 1 Let's consider the 2D frame $\Theta = \{A, B\}$ and two experts providing the following quantitative belief assignments (masses) $m_1(\cdot)$ and $m_2(\cdot)$ as described in Table I.

TABLE I
Quantitative Inputs for Example 1

	A	B	$A \cup B$
$m_1(\cdot)$	1/6	3/6	2/6
$m_2(\cdot)$	4/6	1/6	1/6

The disjunctive operator yields the following result

$$\begin{aligned}
m_\vee(A) &= m_1(A)m_2(A) = (1/6) \cdot (4/6) = 4/36, \\
m_\vee(B) &= m_1(B)m_2(B) = (3/6) \cdot (1/6) = 3/36, \\
m_\vee(A \cup B) &= m_1(A)m_2(B) + m_1(B)m_2(A) \\
&\quad + m_1(A)m_2(A \cup B) + m_2(A)m_1(A \cup B) \\
&\quad + m_1(B)m_2(A \cup B) + m_2(B)m_1(A \cup B) \\
&\quad + m_1(A \cup B)m_2(A \cup B) \\
&= (1/6) \cdot (1/6) + (3/6) \cdot (4/6) \\
&\quad + (1/6) \cdot (1/6) + (4/6) \cdot (2/6) \\
&\quad + (3/6) \cdot (1/6) + (1/6) \cdot (2/6) \\
&\quad + (2/6) \cdot (1/6) \\
&= 29/36,
\end{aligned}$$

while the conjunctive operator yields

$$\begin{aligned}
m_\wedge(A) &= m_1(A)m_2(A) + m_1(A)m_2(A \cup B) \\
&\quad + m_2(A)m_1(A \cup B) \\
&= (1/6) \cdot (4/6) + (1/6) \cdot (1/6) + (4/6) \cdot (2/6) \\
&= 13/36, \\
m_\wedge(B) &= m_1(B)m_2(B) + m_1(B)m_2(A \cup B) \\
&\quad + m_2(B)m_1(A \cup B) \\
&= (3/6) \cdot (1/6) + (3/6) \cdot (1/6) + (1/6) \cdot (2/6) \\
&= 8/36,
\end{aligned}$$

$$m_\wedge(A\cup B) = m_1(A\cup B)m_2(A\cup B) = (2/6)\cdot(1/6) = 2/36,$$

$$\begin{aligned} m_\wedge(A\cap B) &\triangleq m_\wedge(A\cap B) = m_1(A)m_2(B) \\ &\quad + m_2(B)m_1(B) \\ &= (1/6)\cdot(1/6) + (4/6)\cdot(3/6) = 13/36. \end{aligned}$$

- **Dempster's rule** [3]

This combination rule has been initially proposed by Dempster and then used by Shafer in DST framework. We assume (without loss of generality) that the sources of evidence are equally reliable. Otherwise a discounting preprocessing is first applied. It is defined on $G^\Theta = 2^\Theta$ by forcing $m_{\mathrm{DS}}(\emptyset) \triangleq 0$ and $\forall A \in G^\Theta \setminus \{\emptyset\}$ by

$$m_{\mathrm{DS}}(A) = \frac{1}{1-k} m_\wedge(A) = \frac{m_\wedge(A)}{1 - m_\wedge(\emptyset)}. \tag{19}$$

When $k = 1$, this rule cannot be used. Dempster's rule of combination can be directly extended for the combination of N independent and equally reliable sources of evidence and its major interest comes essentially from its commutativity and associativity properties. Dempster's rule corresponds to the normalized conjunctive rule by reassigning the mass of total conflict onto all focal elements through the conjunctive operator. The problem enlightened by the famous Zadeh's example [36] is the repartition of the global conflict. Indeed, consider $\Theta = \{A, B, C\}$ and two experts opinions given by $m_1(A) = 0.9$, $m_1(C) = 0.1$, and $m_2(B) = 0.9$, $m_2(C) = 0.1$, the mass given by Dempster's combination is $m_{\mathrm{DS}}(C) = 1$ which looks very counter-intuitive since it reflects the minority opinion. The generalized Zadeh's example proposed by Smarandache and Dezert in [18], shows that the results obtained by Dempster's rule can moreover become totally independent of the numerical values taken by $m_1(\cdot)$ and $m_2(\cdot)$ which is much more surprising and difficult to accept without reserve for practical fusion applications. To resolve this problem, Smets [26] suggested in his Transferable Belief Model (TBM) framework [28] to consider Θ as an open-world and therefore to use the conjunctive rule instead Dempster's rule at the credal level. At credal level $m_\wedge(\emptyset)$ is interpreted as a non-expected solution. The problem is actually just postponed by Smets at the decision/pignistic level since the normalization (division by $1 - m_\wedge(\emptyset)$) is also required in order to compute the pignistic probabilities of elements of Θ. In other words, the non-normalized version of Dempster's rule corresponds to the Smets' fusion rule in the TBM framework working under an open-world assumption, *i.e.* $m_S(\emptyset) = k = m_\wedge(\emptyset)$ and $\forall A \in G^\Theta \setminus \{\emptyset\}$, $m_S(A) = m_\wedge(A)$.

EXAMPLE 2 Let's consider the 2D frame and quantitative masses as given in example 1 and assume Shafer's model (*i.e.* $A \cap B = \emptyset$), then the conflicting quantitative mass $k = m_\wedge(A\cap B) = 13/36$ is redistributed to the sets A, B, $A\cup B$ proportionally with their $m_\wedge(\cdot)$ masses, *i.e.* $m_\wedge(A) = 13/36$, $m_\wedge(B) = 8/36$ and $m_\wedge(A\cup B) = 2/36$ respectively through Demspter's rule (19). One thus gets

$$\begin{aligned} m_{\mathrm{DS}}(\emptyset) &= 0, \\ m_{\mathrm{DS}}(A) &= (13/36)/(1-(13/36)) = 13/23, \\ m_{\mathrm{DS}}(B) &= (8/36)/(1-(13/36)) = 8/23, \\ m_{\mathrm{DS}}(A\cup B) &= (2/36)/(1-(13/36)) = 2/23. \end{aligned}$$

If one prefers to adopt Smets' TBM approach, at the credal level the empty set is now allowed to have positive mass. In this case, one gets

$$\begin{aligned} m_{\mathrm{TBM}}(\emptyset) &= m_\wedge(A\cap B) = 13/36, \\ m_{\mathrm{TBM}}(A) &= 13/36, \\ m_{\mathrm{TBM}}(B) &= 8/36, \\ m_{\mathrm{TBM}}(A\cup B) &= 2/36. \end{aligned}$$

- **Yager's rule** [32]–[34]

Yager admits that in case of high conflict Dempster's rule provides counter-intuitive results. Thus, k plays the role of an absolute discounting term added to the weight of ignorance. The commutative and quasi-associative[3] Yager's rule is given by $m_Y(\emptyset) = 0$ and $\forall A \in G^\Theta \setminus \{\emptyset\}$ by

$$\begin{aligned} m_Y(A) &= m_\wedge(A), \\ m_Y(\Theta) &= m_\wedge(\Theta) + m_\wedge(\emptyset). \end{aligned} \tag{20}$$

EXAMPLE 3 Let's consider the 2D frame and quantitative masses as given in example 1 and assume Shafer's model (*i.e.* $A \cap B = \emptyset$), then the conflicting quantitative mass $k = m_\wedge(A\cap B) = 13/36$ is transferred to total ignorance $A\cup B$. One thus gets

$$\begin{aligned} m_Y(A) &= 13/36, \\ m_Y(B) &= 8/36, \\ m_Y(A\cup B) &= (2/36) + (13/36) = 15/36. \end{aligned}$$

- **Dubois & Prade's rule** [5]

This rule supposes that the two sources are reliable when they are not in conflict and at least one of them is right when a conflict occurs. Then if one believes that a value is in a set X while the other believes that this value is in a set Y, the truth lies in $X\cap Y$ as long $X\cap Y \neq \emptyset$. If $X\cap Y = \emptyset$, then the truth lies in $X\cup Y$. According to this principle, the commutative and quasi-associative Dubois & Prade hybrid rule of combination, which is a reasonable trade-off between precision and reliability, is defined by $m_{DP}(\emptyset) = 0$ and $\forall A \in G^\Theta \setminus \{\emptyset\}$

[3] Quasi-associativity was defined by Yager in [34], and Smarandache and Dezert in [22].

by

$$m_{DP}(A) = m_{\wedge}(A) + \sum_{\substack{X,Y\in G^{\Theta}\\ X\cup Y=A\\ X\cap Y=\emptyset}} m_1(X)m_2(Y). \qquad (21)$$

In Dubois & Prade's rule, the conflicting information is considered more precisely than in Dempster's or Yager's rules since all partial conflicts involved the total conflict are taken into account separately through (21).

The repartition of the conflict is very important because of the non-idempotency of the rules (except the Denœux' rule [4] that can be applied when the dependency between experts is high) and due to the responses of the experts that can be conflicting. Hence, we have defined the auto-conflict [16] in order to quantify the intrinsic conflict of a mass and the distribution of the conflict according to the number of experts.

EXAMPLE 4 Taking back example 1 and assuming Shafer's model for Θ, the quantitative Dubois & Prade's rule gives the same result as quantitative Yager's rule since the conflicting mass, $m_{\wedge}(A\cap B) = 13/36$, is transferred to $A\cup B$, while the other quantitative masses remain unchanged.

- **Proportional Conflict Redistribution (PCR) rules**

PCR5 *for combining two sources*

Smarandache and Dezert proposed five proportional conflict redistribution (PCR) methods [21], [22] to redistribute the partial conflict on the elements implied in the partial conflict. The most efficient for combining two basic belief assignments $m_1(\cdot)$ and $m_2(\cdot)$ is the PCR5 rule given by $m_{\mathrm{PCR5}}(\emptyset) = 0$ and for all $X \in G^{\Theta}$, $X \neq \emptyset$ by

$$m_{\mathrm{PCR5}}(X) = m_{\wedge}(X) + \sum_{\substack{Y\in G^{\Theta}\\ X\cap Y\equiv\emptyset}} \left(\frac{m_1(X)^2 m_2(Y)}{m_1(X)+m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X)+m_1(Y)} \right), \qquad (22)$$

where $m_{\wedge}(\cdot)$ is the conjunctive rule given by (17).

EXAMPLE 5 Let's consider the 2D frame and quantitative masses as given in Example 1 and assume Shafer's model (*i.e.* $A\cap B = \emptyset$), then the conflicting quantitative mass $k = m_{\wedge}(A\cap B) = 13/36$ is redistributed only to elements involved in conflict, A and B (not to $A\cup B$). We repeat that

$$m_{\wedge}(A\cap B) = m_1(A)m_2(B) + m_2(B)m_1(B) = (1/6)\cdot(1/6) + (4/6)\cdot(3/6) = 13/36.$$

So $(1/6)\cdot(1/6) = 1/36$ is redistributed to A and B proportionally to their quantitative masses assigned by the sources (or experts) $m_1(A) = 1/6$ and $m_2(B) = 1/6$

$$\frac{x_{1,A}}{1/6} = \frac{y_{1,B}}{1/6} = \frac{1/36}{(1/6)+(1/6)} = 1/12,$$

hence

$$x_{1,A} = (1/6)\cdot(1/12) = 1/72,$$

and

$$y_{1,B} = (1/6)\cdot(1/12) = 1/72.$$

Similarly $(4/6)\cdot(3/6) = 12/36$ is redistributed to A and B proportionally to their quantitative masses assigned by the sources (or experts) $m_2(A) = 4/6$ and $m_1(B) = 3/6$

$$\frac{x_{2,A}}{4/6} = \frac{y_{2,B}}{3/6} = \frac{12/36}{(4/6)+(3/6)} = 2/7,$$

hence

$$x_{2,A} = (4/6)\cdot(2/7) = 4/21,$$

and

$$y_{2,B} = (3/6)\cdot(2/7) = 1/7.$$

It is easy to check that

$$x_{1,A} + y_{1,B} + x_{2,A} + y_{2,B} = 13/36 = m_{\wedge}(A\cap B).$$

Summing, we get

$$m_{\mathrm{PCR5}}(A) = (13/36) + (1/72) + (4/21) = 285/504 \simeq 0.57,$$

$$m_{\mathrm{PCR5}}(B) = (8/36) + (1/72) + (1/7) = 191/504 \simeq 0.38,$$

$$m_{\mathrm{PCR5}}(A\cup B) = 2/36 \simeq 0.05,$$

$$m_{\mathrm{PCR5}}(A\cap B = \emptyset) = 0.$$

PCR6 *for combining more than two sources*

A generalization of PCR5 fusion rule for combining altogether more than two experts has been proposed by Smarandache and Dezert in [22]. Recently Martin and Osswald [14], [16] studied and formulated a new version of the PCR5 rule, denoted PCR6, for combining more than two sources, say M sources with $M \geq 2$. Martin and Osswald have shown that PCR6 exhibits a better behavior than PCR5 in specific interesting cases. PCR6 rule is defined as follows: $m_{\mathrm{PCR6}}(\emptyset) = 0$ and for all $X \in G^{\Theta}$, $X \neq \emptyset$,

$$m_{\mathrm{PCR6}}(X) = m_{\wedge}(X) + \sum_{i=1}^{M} m_i(X)^2 \sum_{\substack{\bigcap_{k=1}^{M-1} Y_{\sigma_i(k)}\cap X\equiv\emptyset\\ (Y_{\sigma_i(1)},\ldots,Y_{\sigma_i(M-1)})\in(G^{\Theta})^{M-1}}} \cdot \left(\frac{\prod_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})}{m_i(X) + \sum_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})} \right), \qquad (23)$$

where $Y_j \in G^{\Theta}$ is the response of the expert j, $m_j(Y_j)$ the associated belief function and σ_i counts from 1 to

M avoiding i

$$\begin{aligned} \sigma_i(j) &= j \qquad \text{if} \quad j < i, \\ \sigma_i(j) &= j+1 \qquad \text{if} \quad j \geq i. \end{aligned} \tag{24}$$

The idea is here to redistribute the masses of the focal elements giving a partial conflict proportionally to the initial masses on these elements.

In general, for $M \geq 3$ sources, one calculates the total conflict, which is a sum of products; if each product is formed by factors of masses of distinct hypothesis, then PCR6 coincides with PCR5; if at least a product is formed by at least two factors of masses of same hypotheses, then PCR6 is different from PCR5:

- for example: a product like $m_1(A)m_2(A)m_3(B)$, herein we have two masses of hypothesis A;
- or $m_1(A\cup B)m_2(B\cup C)m_3(B\cup C)m_4(B\cup C)$, herein we have three masses of hypothesis $B\cup C$ from four sources.

EXAMPLE 6 For instance, consider three experts expressing their opinion on $\Theta = \{A,B,C,D\}$ in the Shafer's model as described in Table II.

TABLE II
Quantitative Inputs for Example 6

	A	B	$A\cup C$	$A\cup B\cup C\cup D$
$m_1(\cdot)$	0.7	0	0	0.3
$m_2(\cdot)$	0	0.5	0	0.5
$m_3(\cdot)$	0	0	0.6	0.4

The global conflict is given here by $0.21 + 0.14 + 0.09 = 0.44$, coming from

—A, B and $A\cup C$ for the partial conflict 0.21,

—A, B and $A\cup B\cup C\cup D$ for 0.14,

—and B, $A\cup C$ and $A\cup B\cup C\cup D$ for 0.09.

With the generalized PCR6 rule (23), we obtain:

$$\begin{aligned} m_{\text{PCR6}}(A) &= 0.14 + 0.21 + 0.21\cdot\tfrac{7}{18} + 0.14\cdot\tfrac{7}{16} \\ &\simeq 0.493, \\ m_{\text{PCR6}}(B) &= 0.06 + 0.21\cdot\tfrac{5}{18} + 0.14\cdot\tfrac{5}{16} + 0.09\cdot\tfrac{5}{14} \\ &\simeq 0.194, \\ m_{\text{PCR6}}(A\cup C) &= 0.09 + 0.21\cdot\tfrac{6}{18} + 0.09\cdot\tfrac{6}{14} \simeq 0.199, \\ m_{\text{PCR6}}(A\cup B\cup C\cup D) &= 0.06 + 0.14\cdot\tfrac{4}{16} + 0.09\cdot\tfrac{3}{14} \simeq 0.114. \end{aligned}$$

EXAMPLE 7 Let's consider three sources providing quantitative belief masses only on unions.

The conflict is given here by

$$\begin{aligned} m_{\wedge}(\emptyset) &= m_1(A\cup B)m_2(A\cup C)m_3(B\cup C) \\ &= 0.7\cdot 0.6\cdot 0.5 = 0.21. \end{aligned}$$

TABLE III
Quantitative Inputs for Example 7

	$A\cup B$	$B\cup C$	$A\cup C$	$A\cup B\cup C$
$m_1(\cdot)$	0.7	0	0	0.3
$m_2(\cdot)$	0	0	0.6	0.4
$m_3(\cdot)$	0	0.5	0	0.5

With the generalized PCR rule, *i.e.* PCR6, we obtain

$$\begin{aligned} m_{\text{PCR6}}(A) &= 0.21, \\ m_{\text{PCR6}}(B) &= 0.14, \\ m_{\text{PCR6}}(C) &= 0.09, \\ m_{\text{PCR6}}(A\cup B) &= 0.14 + 0.21.\tfrac{7}{18} \simeq 0.2217, \\ m_{\text{PCR6}}(B\cup C) &= 0.06 + 0.21.\tfrac{5}{18} \simeq 0.1183, \\ m_{\text{PCR6}}(A\cup C) &= 0.09 + 0.21.\tfrac{6}{18} = 0.16, \\ m_{\text{PCR6}}(A\cup B\cup C) &= 0.06. \end{aligned}$$

In the sequel, we use the notation PCR for two and more sources.

4. CLASSICAL QUALITATIVE COMBINATION RULES

The classical qualitative combination rules are direct extensions of classical quantitative rules presented in previous section. Since the formulas of qualitative fusion rules are the same as for quantitative rules, they will be not reported in this section. The main difference between quantitative and qualitative approaches lies in the addition, multiplication and division operators one has to use. For quantitative fusion rules, one uses addition, multiplication and division operators on numbers while for qualitative fusion rules one uses the addition, multiplication and division operators on linguistic labels defined as in Section 2.C1.

EXAMPLE 8 Below is a very simple example used to show how classical qualitative fusion rules work. Let's consider the following set of linguistic labels $L = \{L_{\min} = L_0, L_1, L_2, L_3, L_4, L_5, L_{\max} = L_6\}$ and let's assume Shafer's model for the frame $\Theta = \{A,B\}$ we want to work on. In this example, we consider only two experts providing the qualitative belief assignments (masses) $qm_1(\cdot)$ and $qm_2(\cdot)$ as described in Table IV.

TABLE IV
Qualitative Inputs for Example 8

	A	B	$A\cup B$
$qm_1(\cdot)$	L_1	L_3	L_2
$qm_2(\cdot)$	L_4	L_1	L_1

The qualitative belief assignments $qm_1(\cdot)$ and $qm_2(\cdot)$ have been chosen quasi-normalized since $L_1 + L_3 +$

$L_2 = L_6 = L_{\max}$ and respectively $L_4 + L_1 + L_1 = L_6 = L_{\max}$.

- **Qualitative Conjunctive rule (QCR)**

This rule provides $qm_\wedge(\cdot)$ according following derivations

$$\begin{aligned} qm_\wedge(A) &= qm_1(A)qm_2(A) + qm_1(A)qm_2(A\cup B) \\ &\quad + qm_2(A)qm_1(A\cup B) \\ &= L_1L_4 + L_1L_1 + L_4L_2 = L_{\frac{1\cdot4}{6}} + L_{\frac{1\cdot1}{6}} + L_{\frac{4\cdot2}{6}} \\ &= L_{\frac{4+1+8}{6}} = L_{\frac{13}{6}}, \\ qm_\wedge(B) &= qm_1(B)qm_2(B) + qm_1(B)qm_2(A\cup B) \\ &\quad + qm_2(B)qm_1(A\cup B) \\ &= L_3L_1 + L_3L_1 + L_1L_2 = L_{\frac{3\cdot1}{6}} + L_{\frac{3\cdot1}{6}} + L_{\frac{1\cdot2}{6}} \\ &= L_{\frac{3+3+2}{6}} = L_{\frac{8}{6}}, \\ qm_\wedge(A\cup B) &= qm_1(A\cup B)qm_2(A\cup B) = L_2L_1 = L_{\frac{2\cdot1}{6}} \\ &= L_{\frac{2}{6}}, \\ qm_\wedge(A\cap B) &= qm_1(A)qm_2(B) + qm_2(B)qm_1(B) \\ &= L_1L_1 + L_4L_3 = L_{\frac{1\cdot1}{6}} + L_{\frac{4\cdot3}{6}} = L_{\frac{1+12}{6}} = L_{\frac{13}{6}}. \end{aligned}$$

We see that not approximating the indexes (*i.e.* working with refined labels), the quasi-normalization of the qualitative conjunctive rule is kept

$$L_{\frac{13}{6}} + L_{\frac{8}{6}} + L_{\frac{2}{6}} + L_{\frac{13}{6}} = L_{\frac{36}{6}} = L_6 = L_{\max}.$$

But if we approximate each refined label, we get

$$\begin{aligned} &L_{[\frac{13}{6}]} + L_{[\frac{8}{6}]} + L_{[\frac{2}{6}]} + L_{[\frac{13}{6}]} \\ &\quad = L_2 + L_1 + L_0 + L_2 = L_5 \neq L_6 = L_{\max}. \end{aligned}$$

Let's examine the transfer of the conflicting qualitative mass $qm_\wedge(A\cap B) = qm_\wedge(\emptyset) = L_{\frac{13}{6}}$ to the non-empty sets according to the main following combination rules.

- **Qualitative Dempster's rule (extension of classical numerical DS rule to qualitative masses)**

Assuming Shafer's model for the frame Θ (*i.e.* $A\cap B = \emptyset$) and according to DS rule, the conflicting qualitative mass $qm_\wedge(A\cap B) = L_{\frac{13}{6}}$ is redistributed to the sets A, B, $A\cup B$ proportionally with their $qm_\wedge(\cdot)$ masses $L_{\frac{13}{6}}$, $L_{\frac{8}{6}}$, and $L_{\frac{2}{6}}$ respectively,

$$\begin{aligned} \frac{x_A}{L_{\frac{13}{6}}} = \frac{y_B}{L_{\frac{8}{6}}} = \frac{z_{A\cup B}}{L_{\frac{2}{6}}} &= \frac{L_{\frac{13}{6}}}{L_{\frac{13}{6}} + L_{\frac{8}{6}} + L_{\frac{2}{6}}} \\ &= \frac{L_{\frac{13}{6}}}{L_{\frac{23}{6}}} = L_{(\frac{13}{6}\div\frac{23}{6})\cdot6} = L_{(\frac{13}{23})\cdot6} = L_{\frac{78}{23}}. \end{aligned}$$

Therefore, one gets

$$\begin{aligned} x_A &= L_{\frac{13}{6}}\cdot L_{\frac{78}{23}} = L_{(\frac{13}{6}\cdot\frac{78}{23})\div6} = L_{\frac{169}{138}}, \\ y_B &= L_{\frac{8}{6}}\cdot L_{\frac{78}{23}} = L_{(\frac{8}{6}\cdot\frac{78}{23})\div6} = L_{\frac{104}{138}}, \\ z_{A\cup B} &= L_{\frac{2}{6}}\cdot L_{\frac{78}{23}} = L_{(\frac{2}{6}\cdot\frac{78}{23})\div6} = L_{\frac{26}{138}}. \end{aligned}$$

We can check that the qualitative conflicting mass, $L_{\frac{13}{6}}$, has been proportionally split into three qualitative masses

$$L_{\frac{169}{138}} + L_{\frac{104}{138}} + L_{\frac{26}{138}} = L_{\frac{169+104+26}{138}} = L_{\frac{299}{138}} = L_{\frac{13}{6}}.$$

Thus,

$$\begin{aligned} qm_{\mathrm{DS}}(A) &= L_{\frac{13}{6}} + L_{\frac{169}{138}} = L_{\frac{13}{6}+\frac{169}{138}} = L_{\frac{468}{138}}, \\ qm_{\mathrm{DS}}(B) &= L_{\frac{8}{6}} + L_{\frac{104}{138}} = L_{\frac{8}{6}+\frac{104}{138}} = L_{\frac{288}{138}}, \\ qm_{\mathrm{DS}}(A\cup B) &= L_{\frac{2}{6}} + L_{\frac{26}{138}} = L_{\frac{2}{6}+\frac{26}{138}} = L_{\frac{72}{138}}, \\ qm_{\mathrm{DS}}(A\cap B = \emptyset) &= L_0. \end{aligned}$$

$qm_{\mathrm{DS}}(\cdot)$ is quasi-normalized since:

$$L_{\frac{468}{138}} + L_{\frac{288}{138}} + L_{\frac{72}{138}} = L_{\frac{828}{138}} = L_6 = L_{\max}.$$

If we approximate the linguistic labels $L_{\frac{468}{138}}$, $L_{\frac{288}{138}}$ and $L_{\frac{72}{138}}$ in order to work with original labels in L, still $qm_{\mathrm{DS}}(\cdot)$ remains quasi-normalized since:

$$\begin{aligned} qm_{\mathrm{DS}}(A) &\approx L_{[\frac{468}{138}]} = L_3 \\ qm_{\mathrm{DS}}(B) &\approx L_{[\frac{288}{138}]} = L_2 \\ qm_{\mathrm{DS}}(A\cup B) &\approx L_{[\frac{72}{138}]} = L_1 \end{aligned}$$

and $L_3 + L_2 + L_1 = L_6 = L_{\max}$.

- **Qualitative Yager's rule**

With Yager's rule, the qualitative conflicting mass $L_{\frac{13}{6}}$ is entirely transferred to the total ignorance $A\cup B$. Thus,

$$qm_Y(A\cup B) = L_{\frac{2}{6}} + L_{\frac{13}{6}} = L_{\frac{15}{6}}$$

and

$$qm_Y(A\cap B) = qm_Y(\emptyset) = L_0$$

while the others remain the same

$$\begin{aligned} qm_Y(A) &= L_{\frac{13}{6}}, \\ qm_Y(B) &= L_{\frac{8}{6}}. \end{aligned}$$

$qm_Y(\cdot)$ is quasi-normalized since

$$L_{\frac{13}{6}} + L_{\frac{8}{6}} + L_{\frac{15}{6}} = L_{\frac{36}{6}} = L_6 = L_{\max}.$$

If we approximate the linguistic labels $L_{\frac{13}{6}}$, $L_{\frac{8}{6}}$ and $L_{\frac{15}{6}}$, still $qm_Y(\cdot)$ happens to remain quasi-normalized since

$$\begin{aligned} qm_Y(A) &\approx L_{[\frac{13}{6}]} = L_2, \\ qm_Y(B) &\approx L_{[\frac{8}{6}]} = L_1, \\ qm_Y(A\cup B) &\approx L_{[\frac{15}{6}]} = L_3, \end{aligned}$$

and $L_2 + L_1 + L_3 = L_6 = L_{\max}$.

• **Qualitative Dubois & Prade's rule**

In this example the Qualitative Dubois & Prade's rule gives the same result as qualitative Yager's rule since the conflicting mass, $qm_{\wedge}(A\cap B)=L_{\frac{13}{6}}$, is transferred to $A\cup B$, while the other qualitative masses remain unchanged.

• **Qualitative Smets' TBM rule**

Smets' TBM approach allows keeping mass on the empty set. One gets

$$qm_{\mathrm{TBM}}(A)=L_{\frac{13}{6}},$$

$$qm_{\mathrm{TBM}}(B)=L_{\frac{8}{6}},$$

$$qm_{\mathrm{TBM}}(A\cup B)=L_{\frac{2}{6}},$$

$$qm_{\mathrm{TBM}}(\emptyset)=L_{\frac{13}{6}}.$$

Of course $qm_{\mathrm{TBM}}(\cdot)$ is also quasi-normalized.

However if we approximate, $qm_{\mathrm{TBM}}(\cdot)$ does not remain quasi-normalized in this case since

$$qm_{\mathrm{TBM}}(A)\approx L_{[\frac{13}{6}]}=L_2,$$

$$qm_{\mathrm{TBM}}(B)\approx L_{[\frac{8}{6}]}=L_1,$$

$$qm_{\mathrm{TBM}}(A\cup B)\approx L_{[\frac{2}{6}]}=L_0,$$

$$qm_{\mathrm{TBM}}(\emptyset)\approx L_{[\frac{13}{6}]}=L_2,$$

and $L_2+L_1+L_0+L_2=L_5\neq L_6=L_{\max}$.

• **Qualitative PCR (QPCR)**

The conflicting qualitative mass, $qm_{\wedge}(A\cap B)=L_{\frac{13}{6}}$, is redistributed only to elements involved in conflict, A and B (not to $A\cup B$). We repeat that

$$\begin{aligned}qm_{\mathrm{PCR}}(A\cap B)&=qm_1(A)qm_2(B)+qm_2(B)qm_1(B)\\&=L_1L_1+L_4L_3=L_{\frac{1\cdot1}{6}}+L_{\frac{4\cdot3}{6}}=L_{\frac{1+12}{6}}=L_{\frac{13}{6}}.\end{aligned}$$

So $L_{\frac{1}{6}}$ is redistributed to A and B proportionally to their qualitative masses assigned by the sources (or experts) $qm_1(A)=L_1$ and $qm_2(B)=L_1$

$$\frac{x_{1,A}}{L_1}=\frac{y_{1,B}}{L_1}=\frac{L_{\frac{1}{6}}}{L_1+L_1}=\frac{L_{\frac{1}{6}}}{L_2}=L_{(\frac{1}{6}\div2)\cdot6}=L_{\frac{1}{2}}.$$

Hence

$$x_{1,A}=L_1\cdot L_{\frac{1}{2}}=L_{(1\cdot\frac{1}{2})\div6}=L_{\frac{1}{12}}$$

and

$$y_{1,B}=L_1\cdot L_{\frac{1}{2}}=L_{\frac{1}{12}}.$$

Similarly $L_{\frac{12}{6}}$ is redistributed to A and B proportionally to their qualitative masses assigned by the sources (or experts) $qm_2(A)=L_4$ and $qm_1(B)=L_3$

$$\frac{x_{2,A}}{L_4}=\frac{y_{2,B}}{L_3}=\frac{L_{\frac{12}{6}}}{L_4+L_3}=\frac{L_{\frac{12}{6}}}{L_7}=L_{(\frac{12}{6}\div7)\cdot6}=L_{\frac{12}{7}}.$$

Hence

$$x_{2,A}=L_4\cdot L_{\frac{12}{7}}=L_{(4\cdot\frac{12}{7})\div6}=L_{\frac{8}{7}}$$

and

$$y_{2,B}=L_3\cdot L_{\frac{12}{7}}=L_{(3\cdot\frac{12}{7})\div6}=L_{\frac{6}{7}}.$$

Summing, we get

$$qm_{\mathrm{PCR}}(A)=L_{\frac{13}{6}}+L_{\frac{1}{12}}+L_{\frac{8}{7}}=L_{\frac{285}{84}},$$

$$qm_{\mathrm{PCR}}(B)=L_{\frac{8}{6}}+L_{\frac{1}{12}}+L_{\frac{6}{7}}=L_{\frac{191}{84}},$$

$$qm_{\mathrm{PCR}}(A\cup B)=L_{\frac{2}{6}}=L_{\frac{28}{84}},$$

$$qm_{\mathrm{PCR}}(A\cap B=\emptyset)=L_0.$$

$qm_{\mathrm{PCR}}(\cdot)$ is quasi-normalized since

$$L_{\frac{285}{84}}+L_{\frac{191}{84}}+L_{\frac{28}{84}}=L_{\frac{504}{84}}=L_6=L_{\max}.$$

However, if we approximate, it is not quasi-normalized any longer since

$$L_{[\frac{285}{84}]}+L_{[\frac{191}{84}]}+L_{[\frac{28}{84}]}=L_3+L_2+L_0=L_5\neq L_6=L_{\max}.$$

In general, if we do not approximate, and we work with quasi-normalized qualitative masses, no matter what fusion rule we apply, the result will be quasi-normalized. If we approximate, many times the quasi-normalization is lost.

5. GENERALIZATION OF QUANTITATIVE FUSION RULES

In [1], [29] we can find two propositions of a general formulation of the combination rules. In the first one, Smets considers the combination rules from a matrix notation and find the shape of this matrix according to some assumptions on the rule, such as linearity, commutativity, associativity, etc. In the second one, a generic operator is defined from the plausibility functions.

A general formulation of the global conflict repartition have been proposed in [8], [12] for all $X\in2^{\Theta}$ by

$$m_c(X)=m_{\wedge}(X)+w(X)m_{\wedge}(\emptyset),\tag{25}$$

where $\sum_{X\in2^{\Theta}}w(X)=1$. The problem is the choice of the weights $w(X)$.

A. How to Choose Conjunctive and Disjunctive Rules?

We have seen that conjunctive rule reduces the imprecision and uncertainty but can be used only if one of the experts is reliable, whereas the disjunctive rule can be used when the experts are not reliable, but allows a loss of specificity.

Hence, Florea [7] proposes a weighted sum of these two rules according to the global conflict $k=m_{\wedge}(\emptyset)$

given for $X \in 2^\Theta$ by:

$$m_{\text{Flo}}(X) = \beta_1(k) m_\vee(X) + \beta_2(k) m_\wedge(X), \tag{26}$$

where β_1 and β_2 can admit $k = \frac{1}{2}$ as symmetric weight:

$$\begin{aligned} \beta_1(k) &= \frac{k}{1-k+k^2}, \\ \beta_2(k) &= \frac{1-k}{1-k+k^2}. \end{aligned} \tag{27}$$

Consequently, if the global conflict is high (k near 1) the behavior of this rule will give more importance to the disjunctive rule. Thus, this rule considers the global conflict coming from the non-reliability of the experts.

In order to take into account the weights more precisely in each partial combination, we propose the following new rule. For two basic belief assignments m_1 and m_2 and for all $X \in G^\Theta$, $X \neq \emptyset$ we have:

$$\begin{aligned} m_{\text{Mix}}(X) = &\sum_{Y_1 \cup Y_2 = X} \delta_1(Y_1,Y_2) m_1(Y_1) m_2(Y_2) \\ &+ \sum_{Y_1 \cap Y_2 = X} \delta_2(Y_1,Y_2) m_1(Y_1) m_2(Y_2). \end{aligned} \tag{28}$$

Of course, if $\delta_1(Y_1,Y_2) = \beta_1(k)$ and $\delta_2(Y_1,Y_2) = \beta_2(k)$ we obtain Florea's rule. In the same manner, if $\delta_1(Y_1,Y_2) = 1 - \delta_2(Y_1,Y_2) = 0$ we obtain the conjunctive rule and if $\delta_1(Y_1,Y_2) = 1 - \delta_2(Y_1,Y_2) = 1$ the disjunctive rule. If $\delta_1(Y_1,Y_2) = 1 - \delta_2(Y_1,Y_2) = \mathbb{1}_{Y_1 \cap Y_2 = \emptyset}$ we retrieve Dubois and Prade's rule and the partial conflict can be considered, whereas the rule (26).

The choice of $\delta_1(Y_1,Y_2) = 1 - \delta_2(Y_1,Y_2)$ can be done by a dissimilarity such as:

$$\delta_1(Y_1,Y_2) = \delta(Y_1,Y_2) \triangleq 1 - \frac{\mathcal{C}(Y_1 \cap Y_2)}{\min\{\mathcal{C}(Y_1),\mathcal{C}(Y_2)\}}, \tag{29}$$

or

$$\delta_1(Y_1,Y_2) = \eta(Y_1,Y_2) \triangleq 1 - \frac{\mathcal{C}(Y_1 \cap Y_2)}{\max\{\mathcal{C}(Y_1),\mathcal{C}(Y_2)\}}, \tag{30}$$

where $\mathcal{C}(Y_1)$ is the cardinality of Y_1. In the case of the DST framework, $\mathcal{C}(Y_1)$ is the number of distinct elements of Y_1. In the case of the DSmT, $\mathcal{C}(Y_1)$ is the DSm cardinality given by the number of parts of Y_1 in the Venn diagram of the problem [18]. $\delta(\cdot,\cdot)$ in (29) is actually not a proper dissimilarity measure (*e.g.* $\delta(Y_1,Y_2) = 0$ does not imply $Y_1 = Y_2$), but $\eta(\cdot,\cdot)$ defined in (30) is a proper dissimilarity measure. We can also take for $\delta_2(Y_1,Y_2)$, the Jaccard's distance, *i.e.* $\delta_2(Y_1,Y_2) = d(Y_1,Y_2)$ given by

$$d(Y_1,Y_2) = \frac{\mathcal{C}(Y_1 \cap Y_2)}{\mathcal{C}(Y_1 \cup Y_2)}, \tag{31}$$

used by [10] on the belief functions. Note that d is not a distance in the case of DSmT. Thus, if we have a partial conflict between Y_1 and Y_2, $\mathcal{C}(Y_1 \cap Y_2) = 0$ and the rule transfers the mass on $Y_1 \cup Y_2$. In the case $Y_1 \subset Y_2$ (or the contrary), $Y_1 \cap Y_2 = Y_1$ and $Y_1 \cup Y_2 = Y_2$, so with $\delta_1(\cdot,\cdot) = \delta(\cdot,\cdot)$ the rule transfers the mass on Y_1 and with $\delta_1(\cdot,\cdot) = 1 - d(\cdot,\cdot)$ it transfers the mass on Y_1 and Y_2 according to the ratio $(\mathcal{C}(Y_1)/\mathcal{C}(Y_2))$ of the cardinalities. In the case $Y_1 \cap Y_2 \neq Y_1, Y_2$ and $\emptyset$, the rule transfers the mass on $Y_1 \cap Y_2$ and $Y_1 \cup Y_2$ according to $\delta(\cdot,\cdot)$ and $d(\cdot,\cdot)$.

EXAMPLE 9 (on the derivation of the weights) Let's consider a frame of discernment $\Theta = \{A,B,C\}$ in Shafer's model (*i.e.* all intersections empty).

a) We compute the first similarity weights $\delta_2(\cdot,\cdot) = 1 - \delta(\cdot,\cdot)$ using values presented in Table V.

TABLE V
Values for $1 - \delta(\cdot,\cdot)$

$\delta_2(\cdot,\cdot) = 1 - \delta(\cdot,\cdot)$	A	B	C	$A \cup B$
A	1	0	0	1
B	0	1	0	1
C	0	0	1	0
$A \cup B$	1	1	0	1

We have

$$\delta_2(A,A) = \frac{\mathcal{C}(A \cap A)}{\min\{\mathcal{C}(A),\mathcal{C}(A)\}} = \frac{\mathcal{C}(A)}{\mathcal{C}(A)} = 1,$$

$$\delta_2(A,B) = \frac{\mathcal{C}(A \cap B)}{\min\{\mathcal{C}(A),\mathcal{C}(B)\}} = 0,$$

because $A \cap B = \emptyset$ and $\mathcal{C}(\emptyset) = 0$. Then

$$\delta_2(A,A \cup B) = \frac{\mathcal{C}(A \cap (A \cup B))}{\min\{\mathcal{C}(A),\mathcal{C}(A \cup B)\}} = \frac{\mathcal{C}(A)}{\mathcal{C}(A)} = 1,$$

etc.

Whence, the first dissimilarity weights $\delta_1(\cdot,\cdot)$ defined by (29), *i.e.* $\delta_1(X,Y) = 1 - \delta_2(X,Y)$ take the values as presented in Table VI.

TABLE VI
Values for $\delta(\cdot,\cdot)$

$\delta_1(\cdot,\cdot) = \delta(\cdot,\cdot)$	A	B	C	$A \cup B$
A	0	1	1	0
B	1	0	1	0
C	1	1	0	1
$A \cup B$	0	0	1	0

The first similarity and dissimilarity weights $\delta_2(\cdot,\cdot)$ and $\delta_1(\cdot,\cdot)$ are not quite accurate, since for example: $\delta_2(A,A \cup B) = 1$, *i.e.* A and $A \cup B$ are 100% similar (which is not the case since $A \neq A \cup B$) and $\delta_1(A,A \cup B) = 1 - \delta_2(A,A \cup B) = 1 - 1 = 0$, *i.e.* A and $A \cup B$ are 100% dissimilar (which is not the case either since $A \cap (A \cup B) \neq \emptyset$).

b) The second similarity weights $\delta_2(\cdot,\cdot) = 1 - \eta(\cdot,\cdot)$ given by (30) overcomes this problem. We obtain on the previous example with values given in Table VII.

TABLE VII
Values for $1 - \eta(\cdot,\cdot)$

$\delta_2(\cdot,\cdot) = 1 - \eta(\cdot,\cdot)$	A	B	C	$A \cup B$
A	1	0	0	1/2
B	0	1	0	1/2
C	0	0	1	0
$A \cup B$	1/2	1/2	0	1

Then,

$$\delta_2(A,A) = 1 - \eta(A,A) = \frac{\mathcal{C}(A \cap A)}{\max\{\mathcal{C}(A),\mathcal{C}(A)\}} = \frac{\mathcal{C}(A)}{\mathcal{C}(A)} = 1,$$

$$\delta_2(A,B) = 1 - \eta(A,B) = \frac{\mathcal{C}(A \cap B)}{\max\{\mathcal{C}(A),\mathcal{C}(B)\}} = 0,$$

because $A \cap B = \emptyset$ and $\mathcal{C}(\emptyset) = 0$. Hence,

$$\delta_2(A,A \cup B) = 1 - \eta(A,A \cup B) = \frac{\mathcal{C}(A \cap (A \cup B))}{\max\{\mathcal{C}(A),\mathcal{C}(A \cup B)\}} = \frac{\mathcal{C}(A)}{\mathcal{C}(A \cup B)} = \frac{1}{2}$$

which is better than $\delta_2(A,A \cup B) = 1 - \delta(A,A \cup B) = 1$. etc.

Whence, the second dissimilarity weights $\eta(\cdot,\cdot)$ take the values presented in Table VIII.

TABLE VIII
Values for $\eta(\cdot,\cdot)$

$\delta_1(\cdot,\cdot) = \eta(\cdot,\cdot)$	A	B	C	$A \cup B$
A	0	1	1	1/2
B	1	0	1	1/2
C	1	1	0	1
$A \cup B$	1/2	1/2	1	0

Then, $\eta(A,A \cup B) = 1 - \frac{1}{2} = \frac{1}{2}$, which is better than $\delta_1(A,A \cup B) = \delta(A,A \cup B) = 0$.

The second similarity weight coincides with Jaccard's distance in Shafer's model, but in hybrid and free models, they are generally different. Hence if we consider a Shafer's model, one gets for all Y_1, Y_2 in G^Θ

$$d(Y_1,Y_2) = 1 - \eta(Y_1,Y_2).$$

Smarandache defined in [23] the degree of intersection of two sets as Jaccard's distance, and also the degree of union of two sets, and the degree of inclusion of a set into another set and improved many fusion rules by inserting these degrees in the fusion rules' formulas.

EXAMPLE 10 (with Shafer's model) Consider the following example for two (quantitative) experts providing $m_1(\cdot)$ and $m_2(\cdot)$ on $\Theta = \{A,B,C\}$ and let's assume that Shafer's model holds (*i.e.* A, B and C are truly exclusive). Consider the following example given by Table IX for two (quantitative) experts.

TABLE IX
Quantitative Inputs and Fusion Result

	$m_1(\cdot)$	$m_2(\cdot)$	$m_\wedge$	$m_{\text{Mix},\delta}$	$m_{\text{Mix},\eta}$ $m_{\text{Mix},d}$
$\emptyset$	0	0	0.2	0	0
A	0.3	0	0.3	0.24	0.115
B	0	0.2	0.14	0.14	0.06
$A \cup B$	0.4	0	0.12	0.18	0.18
C	0	0.2	0.06	0.06	0.02
$A \cup C$	0	0.3	0.09	0.15	0.165
$A \cup B \cup C$	0.3	0.3	0.09	0.23	0.46

When taking $\delta_1(\cdot,\cdot) = \delta(\cdot,\cdot)$ according to (29), one obtains the results given in Table X.

TABLE X
Values for $\delta(\cdot,\cdot)$

$\delta_1(\cdot,\cdot) = \delta(\cdot,\cdot)$	A	$A \cup B$	$A \cup B \cup C$
B	1	0	0
C	1	1	0
$A \cup C$	0	1/2	0
$A \cup B \cup C$	0	0	0

where the columns are the focal elements of the basic belief assignment given by the expert 1 and the rows are the focal elements of the basic belief assignment given by expert 2. The mass 0.2 on $\emptyset$ come from the responses A and C with a value of 0.06, from the responses A and B with a value of 0.06 and from the responses $A \cup B$ and C with a value of 0.08. These three values are transferred respectively on $A \cup C$, $A \cup B$ and $A \cup B \cup C$. The mass 0.12 on A given by the responses $A \cup B$ and $A \cup C$ is transferred on A with a value of 0.06 and on $A \cup B \cup C$ with the same value.

When taking $\delta_1(\cdot,\cdot) = \eta(\cdot,\cdot)$ or $\delta_1(\cdot,\cdot) = 1 - d(\cdot,\cdot)$ according to (30) and (31), one obtains the results presented in Table XI.

TABLE XI
Values for $\eta(\cdot,\cdot)$ or $1 - d(\cdot,\cdot)$

$\delta_1(\cdot,\cdot) = \eta(\cdot,\cdot)$ $\delta_1(\cdot,\cdot) = 1 - d(\cdot,\cdot)$	A	$A \cup B$	$A \cup B \cup C$
B	1	1/2	2/3
C	1	1	2/3
$A \cup C$	1/2	2/3	1/3
$A \cup B \cup C$	2/3	1/3	0

With $\delta_1(\cdot,\cdot) = \eta$ or $\delta_1(\cdot,\cdot) = 1 - d(\cdot,\cdot)$, the rule is more disjunctive: more masses are transferred on the ignorance.

Note that $\delta_1(\cdot,\cdot) = \delta(\cdot,\cdot)$ can be used when the experts are considered reliable. In this case, we consider the most precise response. With $\delta_1(\cdot,\cdot) = \eta(\cdot,\cdot)$ or $\delta_1(\cdot,\cdot) = 1 - d(\cdot,\cdot)$, we get the conjunctive rule only when the experts provide the same response, otherwise we consider the doubtful responses and we transfer the masses in proportion of the imprecision of the responses (given by the cardinality of the responses) on the part in agreement and on the partial ignorance.

EXAMPLE 11 (with a hybrid model) Consider the same example with two (quantitative) experts providing $m_1(\cdot)$ and $m_2(\cdot)$ on the frame of discernment $\Theta = \{A, B, C\}$ with the following integrity constraints: $A \cap B \neq \emptyset$, $A \cap C = \emptyset$ and $B \cap C = \emptyset$ (which defines a so-called DSm-hybrid model [18]). The results are given in Table XII.

TABLE XII
Quantitative Inputs and Fusion Result

	$m_1(\cdot)$	$m_2(\cdot)$	$m_\wedge$	$m_{\text{Mix},\delta}$	$m_{\text{Mix},\eta}$	$m_{\text{Mix},d}$
$\emptyset$	0	0	0.14	0	0	0
$A \cap B$	0	0	0.06	0.03	0.03	0.02
A	0.3	0	0.3	0.26	0.205	0.185
B	0	0.2	0.14	0.14	0.084	0.084
$A \cup B$	0.4	0	0.12	0.15	0.146	0.156
C	0	0.2	0.06	0.06	0.015	0.015
$A \cup C$	0	0.3	0.09	0.15	0.1575	0.1575
$A \cup B \cup C$	0.3	0.3	0.09	0.21	0.3625	0.3825

When taking $\delta_1(\cdot,\cdot) = \delta(\cdot,\cdot)$ according to (29), one obtains results presented in Table XIII.

TABLE XIII
Values for $\delta(\cdot,\cdot)$

$\delta_1(\cdot,\cdot) = \delta(\cdot,\cdot)$	A	$A \cup B$	$A \cup B \cup C$
B	1/2	0	0
C	1	1	0
$A \cup C$	0	1/3	0
$A \cup B \cup C$	0	0	0

When taking $\delta_1(\cdot,\cdot) = \eta(\cdot,\cdot)$ according to (30), one obtains results presented in Table XIV.

TABLE XIV
Values for $\eta(\cdot,\cdot)$

$\delta_1(\cdot,\cdot) = \eta(\cdot,\cdot)$	A	$A \cup B$	$A \cup B \cup C$
B	1/2	1/3	1/2
C	1	1	3/4
$A \cup C$	1/3	1/3	1/4
$A \cup B \cup C$	1/2	1/4	0

When taking $\delta_1(\cdot,\cdot) = 1 - d(\cdot,\cdot)$ according to (31), one obtains results presented in Table XV.

TABLE XV
Values for $1 - d(\cdot,\cdot)$

$\delta_1(\cdot,\cdot) = 1 - d(\cdot,\cdot)$	A	$A \cup B$	$A \cup B \cup C$
B	2/3	1/3	1/2
C	1	1	3/4
$A \cup C$	1/3	1/2	1/4
$A \cup B \cup C$	1/2	1/4	0

For more than two experts, say $M > 2$, if the intersection of the responses of the M experts is not empty, we can still transfer on the intersection and the union, and (29) and (30) become

$$\delta_1(Y_1,\ldots,Y_M) = \delta(Y_1,\ldots,Y_M) = 1 - \frac{\mathcal{C}(Y_1 \cap \cdots \cap Y_M)}{\min_{1\le i\le M} \mathcal{C}(Y_i)}, \tag{32}$$

and

$$\delta_1(Y_1,\ldots,Y_M) = \eta(Y_1,\ldots,Y_M) = 1 - \frac{\mathcal{C}(Y_1 \cap \cdots \cap Y_M)}{\max_{1\le i\le M} \mathcal{C}(Y_i)}. \tag{33}$$

From (31), we can define δ_1 by:

$$\delta_1(Y_1,\ldots,Y_M) = 1 - \frac{\mathcal{C}(Y_1 \cap \cdots \cap Y_M)}{\mathcal{C}(Y_1 \cup \cdots \cup Y_M)}. \tag{34}$$

Finally, the mixed rule for $M \geq 2$ experts is given by:

$$m_{\text{Mix}}(X) = \sum_{Y_1 \cup \cdots \cup Y_M = X} \delta_1(Y_1,\ldots,Y_M) \prod_{j=1}^{M} m_j(Y_j) + \sum_{Y_1 \cap \cdots \cap Y_M = X} (1 - \delta_1(Y_1,\ldots,Y_M)) \prod_{j=1}^{M} m_j(Y_j). \tag{35}$$

This formulation can be interesting according to the coherence of the responses of the experts. However, it does not allow the repartition of the partial conflict in an other way than the Dubois and Prade's rule.

B. A Discounting Proportional Conflict Repartition Rule

The PCR6 redistributes the masses of the conflicting focal elements proportionally to the initial masses on these elements. First, the repartition concerns only on the elements involved in the partial conflict. We can apply a discounting procedure in the combination rule in order to transfer a part of the partial conflict on the partial ignorance. This new discounting PCR (noted DPCR) can be expressed for two basic belief assign-

ments $m_1(\cdot)$ and $m_2(\cdot)$ and for all $X \in G^\Theta$, $X \neq \emptyset$ by

$$m_{\mathrm{DPCR}}(X) = m_\wedge(X) + \sum_{\substack{Y \in G^\Theta \\ X \cap Y \equiv \emptyset}} \alpha \cdot \left(\frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right) + \sum_{\substack{Y_1 \cup Y_2 = X \\ Y_1 \cap Y_2 \equiv \emptyset}} (1-\alpha) \cdot m_1(Y_1) m_2(Y_2), \quad (36)$$

where $\alpha \in [0,1]$ is the discounting factor. Note that we can also apply a discounting procedure on the masses before the combination as shown in (7). Here the discounting factor is introduced in order to transfer a part of the partial conflict on partial ignorance. We propose in (39) and (40) different ways for choosing this factor α.

Hence, DPCR fusion rule is a combination of PCR and Dubois-Prade (or DSmH[4]) rules. In an analogue way we can combine other fusion rules, two or more in the same formula, getting new mixed formulas. So that in a general case, for $M \geq 2$ experts, we can extend the previous rule as

$$m_{\mathrm{DPCR}}(X) = m_\wedge(X) + \sum_{i=1}^{M} m_i(X)^2 \sum_{\substack{\bigcap_{k=1}^{M-1} Y_{\sigma_i(k)} \cap X = \emptyset \\ (Y_{\sigma_i(1)}, \ldots, Y_{\sigma_i(M-1)}) \in (G^\Theta)^{M-1}}} \cdot \alpha \cdot \left(\frac{\prod_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})}{m_i(X) + \sum_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})} \right) + \sum_{\substack{Y_1 \cup \cdots \cup Y_M = X \\ Y_1 \cap \cdots \cap Y_M \equiv \emptyset}} (1-\alpha) \cdot \prod_{j=1}^{M} m_j(Y_j), \quad (37)$$

where $Y_j \in G^\Theta$ is a response of the expert j, $m_j(Y_j)$ its assigned mass and σ_i is given by (24).

Hence, if we choose as discounting factor $\alpha = 0.9$ in the previous example, we obtain

$$\begin{aligned} m_{\mathrm{DPCR}}(A) &= 0.14 + 0.21 + 0.21 \cdot \tfrac{7}{18} \cdot 0.9 + 0.14 \cdot \tfrac{7}{16} \cdot 0.9 \simeq 0.479, \\ m_{\mathrm{DPCR}}(B) &= 0.06 + 0.21 \cdot \tfrac{5}{18} \cdot 0.9 + 0.14 \cdot \tfrac{5}{16} \cdot 0.9 + 0.09 \cdot \tfrac{5}{14} \cdot 0.9 \simeq 0.181, \\ m_{\mathrm{DPCR}}(A \cup C) &= 0.09 + 0.21 \cdot \tfrac{6}{18} \cdot 0.9 + 0.09 \cdot \tfrac{6}{14} \cdot 0.9 \simeq 0.187, \\ m_{\mathrm{DPCR}}(A \cup B \cup C) &= 0.21 \cdot 0.1 = 0.021, \\ m_{\mathrm{DPCR}}(A \cup B \cup C \cup D) &= 0.06 + 0.14 \cdot \tfrac{4}{16} \cdot 0.9 + 0.09 \cdot \tfrac{3}{14} \cdot 0.9 + 0.14 \cdot 0.1 + 0.09 \cdot 0.1 \simeq 0.132. \end{aligned}$$

However, in this example, the partial conflict due to the experts 1, 2 and 3 saying A, B, and $A \cup C$ respectively, the conflict is 0.21. Nonetheless, only the experts 1 and 2 and the experts 2 and 3 are in conflict. The experts 1 and 3 are not in conflict.

Now, consider another case where the experts 1, 2 and 3 say A, B, and C respectively with the same conflict 0.21. In both cases, the DPCR rule transfers the masses with the same weight α. Although, we could prefer transfer more mass on Θ in the second than in the first case.

Consequently, the transfer of mass can depend on the existence of conflict between each pair of experts. We define the conflict function giving the number of experts in conflict two by two for each response $Y_i \in G^\Theta$ of the expert i as the number of responses of the other experts in conflict with i. A function f_i is defined by the mapping of $(G^\Theta)^M$ onto $[0, 1/M]$ with

$$f_i(Y_1, \ldots, Y_M) = \frac{\sum_{j=1}^{M} \mathbb{1}_{\{Y_j \cap Y_i = \emptyset\}}}{M(M-1)}. \quad (38)$$

Hence, we can choose α depending on the response of the experts such as

$$\alpha(Y_1, \ldots, Y_M) = 1 - \sum_{i=1}^{M} f_i(Y_1, \ldots, Y_M). \quad (39)$$

In this case $\alpha \in [0,1]$, we do not transfer the mass on elements that can be written as the union of the responses of the experts.

Therefore, if we consider again our previous example we obtain

$$\begin{aligned} \alpha(A, B, A \cup C) &= 1 - \tfrac{2}{3} = \tfrac{1}{3}, \\ \alpha(A, B, A \cup B \cup C \cup D) &= 1 - \tfrac{1}{3} = \tfrac{2}{3}, \\ \alpha(A \cup B \cup C \cup D, B, A \cup C) &= 1 - \tfrac{1}{3} = \tfrac{2}{3}. \end{aligned}$$

Thus the provided mass by the DPCR is

$$\begin{aligned} m_{\mathrm{DPCR}}(A) &= 0.14 + 0.21 + 0.21 \cdot \tfrac{7}{18} \cdot \tfrac{1}{3} + 0.14 \cdot \tfrac{7}{16} \cdot \tfrac{2}{3} \simeq 0.418, \\ m_{\mathrm{DPCR}}(B) &= 0.06 + 0.21 \cdot \tfrac{5}{18} \cdot \tfrac{1}{3} + 0.14 \cdot \tfrac{5}{16} \cdot \tfrac{2}{3} + 0.09 \cdot \tfrac{5}{14} \cdot \tfrac{2}{3} \simeq 0.130, \end{aligned}$$

[4]The DSmH rule is an extension of Dubois-Prade's rule which has been proposed in the DSmT framework in order to work with hybrid models including non-existential constraints. See [18] for details and examples.

$$m_{\mathrm{DPCR}}(A\cup C) = 0.09 + 0.21\cdot\tfrac{6}{18}\cdot\tfrac{1}{3} + 0.09\cdot\tfrac{6}{14}\cdot\tfrac{2}{3}$$
$$\simeq 0.139,$$
$$m_{\mathrm{DPCR}}(A\cup B\cup C) = 0.21\cdot\tfrac{2}{3} = 0.140,$$
$$m_{\mathrm{DPCR}}(A\cup B\cup C\cup D) = 0.06 + 0.14\cdot\tfrac{4}{16}\cdot\tfrac{2}{3} + 0.09\cdot\tfrac{3}{14}\cdot\tfrac{2}{3} + 0.14\cdot\tfrac{1}{3} + 0.09\cdot\tfrac{1}{3} \simeq 0.173.$$

We want to take account of the degree of conflict (or non-conflict) within each pair of expert differently for each element. We can consider the non-conflict function given for each expert i by the number of experts not in conflict with i. Hence, we can choose $\alpha_i(Y_1,\ldots,Y_M)$ defined by the mapping of $(G^\Theta)^M$ onto $[0,1/M]$ with

$$\alpha_i(Y_1,\ldots,Y_M) = \frac{1}{M} - f_i(Y_1,\ldots,Y_M) = \frac{\sum_{j=1,j\neq i}^{M}\mathbb{1}_{\{Y_j\cap Y_i\neq\emptyset\}}}{M(M-1)}. \tag{40}$$

The discounting PCR rule (37) can be written for M experts, for all $X\in G^\Theta$, $X\neq\emptyset$ as:

$$m_{\mathrm{DPCR}}(X) = m_\wedge(X) + \sum_{i=1}^{M} m_i(X)^2 \sum_{\substack{\bigcap_{k=1}^{M-1} Y_{\sigma_i(k)}\cap X=\emptyset \\ (Y_{\sigma_i(1)},\ldots,Y_{\sigma_i(M-1)})\in(G^\Theta)^{M-1}}} \cdot\alpha_i\lambda\left(\frac{\prod_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})}{m_i(X)+\sum_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})}\right) + \sum_{\substack{Y_1\cup\cdots\cup Y_M=X \\ Y_1\cap\cdots\cap Y_M=\emptyset}} \left(1-\sum_{i=1}^{M}\alpha_i\right)\prod_{j=1}^{M} m_j(Y_j), \tag{41}$$

where $\alpha_i(X,Y_{\sigma_i(1)},\ldots,Y_{\sigma_i(M-1)})$ is noted α_i for notations convenience and λ depending on $(X,Y_{\sigma_i(1)},\ldots,Y_{\sigma_i(M-1)})$, is chosen to obtain the normalization given by (2). λ is given when $\alpha_i\neq 0$, $\forall i\in\{1,\ldots,M\}$ by:

$$\lambda = \frac{\sum_{i=1}^{M}\alpha_i}{\langle\boldsymbol{\alpha},\boldsymbol{\gamma}\rangle}, \tag{42}$$

where $\langle\boldsymbol{\alpha},\boldsymbol{\gamma}\rangle$ is the scalar product of $\boldsymbol{\alpha}=(\alpha_i)_{i\in\{1,\ldots,M\}}$ and $\boldsymbol{\gamma}=(\gamma_i)_{i\in\{1,\ldots,M\}}$ with:

$$\gamma_i = \frac{m_i(X)}{m_i(X)+\sum_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})}, \tag{43}$$

where $\gamma_i(X,Y_{\sigma_i(1)},\ldots,Y_{\sigma_i(M-1)})$ is noted γ_i for notations convenience.

With this last version of the rule, for α_i given by (40), we obtain on our illustrative example $\lambda=\frac{36}{13}$ when the experts 1, 2 and 3 say A, B, and $A\cup C$ respectively (the conflict is 0.21), $\lambda=\frac{16}{5}$ when the conflict is 0.14 and $\lambda=\frac{56}{17}$ when the conflict is 0.09. Thus, the masses are given by:

$$m_{\mathrm{DPCR}}(A) = 0.14 + 0.21 + 0.21\cdot\tfrac{7}{18}\cdot\tfrac{1}{6}\cdot\tfrac{36}{13} + 0.14\cdot\tfrac{7}{16}\cdot\tfrac{1}{6}\cdot\tfrac{16}{5} \simeq 0.420$$
$$m_{\mathrm{DPCR}}(B) = 0.06 + 0.14\cdot\tfrac{5}{16}\cdot\tfrac{1}{6}\cdot\tfrac{16}{5} + 0.09\cdot\tfrac{5}{14}\cdot\tfrac{1}{6}\cdot\tfrac{56}{17} \simeq 0.101$$
$$m_{\mathrm{DPCR}}(A\cup C) = 0.09 + 0.21\cdot\tfrac{6}{18}\cdot\tfrac{1}{6}\cdot\tfrac{36}{13} + 0.09\cdot\tfrac{6}{14}\cdot\tfrac{1}{6}\cdot\tfrac{56}{17} \simeq 0.143$$
$$m_{\mathrm{DPCR}}(A\cup B\cup C) = 0.21\cdot\tfrac{2}{3} = 0.14$$
$$m_{\mathrm{DPCR}}(A\cup B\cup C\cup D) = 0.06 + 0.14\cdot\tfrac{4}{16}\cdot\tfrac{1}{3}\cdot\tfrac{16}{5} + 0.09\cdot\tfrac{3}{14}\cdot\tfrac{1}{3}\cdot\tfrac{56}{17} + 0.14\cdot\tfrac{1}{3} + 0.09\cdot\tfrac{1}{3} \simeq 0.196.$$

This last rule of combination allows one to consider a "kind of degree" of conflict (a degree of pair of non-conflict), but this degree is not so easy to introduce in the combination rule.

C. A Mixed Discounting Conflict Repartition Rule

In this section, we propose a combination of the mixed rule (35) with the discounting PCR (37). This new *mixed discounting conflict repartition rule* (MDPCR for short) for two quantitative basic belief assignments $m_1(\cdot)$ and $m_2(\cdot)$ is defined by $m_{\mathrm{MDPCR}}(\emptyset)=0$ and for all $X\in G^\Theta$, $X\neq\emptyset$ by:

$$m_{\mathrm{MDPCR}}(X) = \sum_{\substack{Y_1\cup Y_2=X, \\ Y_1\cap Y_2\neq\emptyset}} \delta_1(Y_1,Y_2)\cdot m_1(Y_1)m_2(Y_2) + \sum_{\substack{Y_1\cap Y_2=X, \\ Y_1\cap Y_2\neq\emptyset}} (1-\delta_1(Y_1,Y_2))\cdot m_1(Y_1)m_2(Y_2) + \sum_{\substack{Y\in G^\Theta, \\ X\cap Y=\emptyset}} \alpha\cdot\left(\frac{m_1(X)^2 m_2(Y)}{m_1(X)+m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X)+m_1(Y)}\right) + \sum_{\substack{Y_1\cup Y_2=X, \\ Y_1\cap Y_2=\emptyset}} (1-\alpha)\cdot m_1(Y_1)m_2(Y_2). \tag{44}$$

α can be given by (39) and $\delta_1(\cdot,\cdot)$ by (32) or (34). The weights must be taken in order to get a kind of continuity between the mixed and DPCR rules. In actuality, when the intersection of the responses is almost empty (but not empty) we use the mixed rule, and when this intersection is empty we chose the DPCR rule. In the first case, all the mass is transferred on the union, and in the second case it will be the same according to the partial conflict. Indeed, $\alpha=0$ if the intersection is not empty and $\delta_1=1$ if the intersection is empty. We can also introduce α_i given by (40), and this continuity is conserved.

This rule is given in a general case for M experts, by $m_{\text{MDPCR}}(\emptyset) = 0$ and for all $X \in G^\Theta$, $X \neq \emptyset$ by:

$$
\begin{aligned}
m_{\text{MDPCR}}(X) = & \sum_{\substack{Y_1 \cup \cdots \cup Y_M = X, \\ Y_1 \cap \cdots \cap Y_M \neq \emptyset}} \delta_1(Y_1, \ldots, Y_M) \cdot \prod_{j=1}^{M} m_j(Y_j) \\
& + \sum_{\substack{Y_1 \cap \cdots \cap Y_M = X, \\ Y_1 \cap \cdots \cap Y_M \neq \emptyset}} (1 - \delta_1(Y_1, \ldots, Y_M)) \cdot \prod_{j=1}^{M} m_j(Y_j) \\
& + \sum_{i=1}^{M} m_i(X)^2 \sum_{\substack{\bigcap_{k=1}^{M-1} Y_{\sigma_i(k)} \cap X = \emptyset \\ (Y_{\sigma_i(1)}, \ldots, Y_{\sigma_i(M-1)}) \in (G^\Theta)^{M-1}}} \\
& \cdot \alpha \cdot \left(\frac{\prod_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})}{m_i(X) + \sum_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})} \right) \\
& + \sum_{\substack{Y_1 \cup \cdots \cup Y_M = X, \\ Y_1 \cap \cdots \cap Y_M = \emptyset}} (1 - \alpha) \cdot \prod_{j=1}^{M} m_j(Y_j), \qquad (45)
\end{aligned}
$$

where $Y_j \in G^\Theta$ is the response of the expert j, $m_j(Y_j)$ the associated belief function and σ_i is given by (24). This formula could seem difficult to understand, but it can be implemented easily as shown in [15].

If we take again the previous example, with $\delta_1(\cdot,\cdot)$ given by (32), there is no difference with the DPCR. If $\delta_1(\cdot,\cdot)$ is calculated by (34), the only difference pertains to the mass 0.09 coming from the responses of the three experts: $A \cup B \cup C \cup D$, $A \cup B \cup C \cup D$ and $A \cup C$. This mass is transferred on $A \cup C$ (0.06) and on $A \cup B \cup C \cup D$ (0.03).

The rules presented in the previous section, propose a repartition of the masses giving a partial conflict only (when at most two experts are in discord) and do not take heed of the level of imprecision of the responses of the experts (the non-specificity of the responses). The imprecision of the responses of each expert is only considered by the mixed and MDPCR rules when there is no conflict between the experts. To try to overcome these problems Martin and Osswald have proposed a begin of solutions toward a more general rule [15].

6. GENERALIZATION OF QUALITATIVE FUSION RULES

This section provides two simple examples to show in detail how to extend the generalized quantitative fusion rules proposed in the previous section (*i.e.* the Mixed, the Discounted, and the Mixed Discounted fusion rules) to their qualitative counterparts using our operators on linguistic labels defined in Section 2.C1.

EXAMPLE 12 Fusion of two sources Consider a set of labels $L = \{L_{\min} = L_0, L_1, L_2, L_3, L_4, L_5, L_{\max} = L_6\}$, and a frame of discernment $\Theta = \{A, B, C\}$ in Shafer's model (*i.e.* all intersections empty). Consider the two following qualitative sources of evidence described in Table XVI.

TABLE XVI
Qualitative Inputs for Example 12

	A	B	C	$A \cup B$
$qm_1(\cdot)$	L_2	L_0	L_0	L_4
$qm_2(\cdot)$	L_3	L_2	L_1	L_0

Now let's apply the qualitative versions of Mixed, Discounted, and Mixed Discounted Quantitative Fusion rules (28), (36) and (44) respectively.

- **Qualitative Mixed Dubois-Prade's rule**
  From (28) and Table VI, one gets:

$$
\begin{aligned}
qm^{\delta}_{\text{Mix}}(A) = & \, \delta(A,A) qm_1(A) qm_2(A) \\
& + (1 - \delta(A,A)) qm_1(A) qm_2(A) \\
& + (1 - \delta(A, A \cup B)) qm_1(A) qm_2(A \cup B) \\
& + (1 - \delta(A \cup B, A)) qm_1(A \cup B) qm_2(A) \\
= & \, 0 \cdot L_2 L_3 + 1 \cdot L_2 L_3 + 1 \cdot L_2 L_0 + 1 \cdot L_4 L_3 \\
= & \, L_0 + L_{\frac{2\cdot 3}{6}} + L_{\frac{2\cdot 0}{6}} + L_{\frac{4\cdot 3}{6}} = L_{\frac{6}{6}} + L_{\frac{12}{6}} = L_{\frac{18}{6}}.
\end{aligned}
$$

Similarly, $qm^{\delta}_{\text{Mix}}(B) = L_{\frac{8}{6}}$ and

$$
\begin{aligned}
qm^{\delta}_{\text{Mix}}(C) = & \, \delta(C,C) qm_1(C) qm_2(C) \\
& + (1 - \delta(C,C)) qm_1(C) qm_2(C) \\
= & \, 0 \cdot L_0 L_1 + 1 \cdot L_0 L_1 = L_0,
\end{aligned}
$$

$$
\begin{aligned}
qm^{\delta}_{\text{Mix}}(A \cup B) = & \, \delta(A \cup B, A \cup B) qm_1(A \cup B) qm_2(A \cup B) \\
& + \delta(A, A \cup B) qm_1(A) qm_2(A \cup B) \\
& + \delta(A \cup B, A) qm_1(A \cup B) qm_2(A) \\
& + \delta(B, A \cup B) qm_1(B) qm_2(A \cup B) \\
& + \delta(A \cup B, B) qm_1(A \cup B) qm_2(B) \\
& + \delta(A, B) qm_1(A) qm_2(B) \\
& + \delta(B, A) qm_1(B) qm_2(A) \\
& + (1 - \delta(A \cup B, A \cup B)) qm_1(A \cup B) qm_2(A \cup B) \\
= & \, L_0 + L_0 + L_0 + L_0 + L_0 + 1 \cdot L_2 L_2 + 1 \cdot L_0 L_3 \\
= & \, L_{\frac{2\cdot 2}{6}} + L_{\frac{0\cdot 3}{6}} = L_{\frac{4}{6}}.
\end{aligned}
$$

Note: The first five terms of previous sum take value L_0 since $\delta_1(\cdot,\cdot) = 0$ for each of them. Then

$$
\begin{aligned}
qm^{\delta}_{\text{Mix}}(A \cup C) = & \, \delta(A,C) qm_1(A) qm_2(C) \\
& + (1 - \delta(C,A)) qm_1(C) qm_2(A) \\
= & \, 1 \cdot L_2 L_1 + 1 \cdot L_0 L_3 = L_{\frac{2\cdot 1}{6}} + L_{\frac{0\cdot 3}{6}} = L_{\frac{2}{6}},
\end{aligned}
$$

$$qm^{\delta}_{\text{Mix}}(A\cup B\cup C) = \delta(C, A\cup B)qm_1(C)qm_2(A\cup B) + \delta(A\cup B, C)qm_1(A\cup B)qm_2(C) = 1\cdot L_0L_0 + 1\cdot L_4L_1 = L_{\frac{4}{6}}.$$

This coincides with normal qualitative Dubois-Prade's and DSmH fusion rules. $qm^{\delta}_{\text{Mix}}(\cdot)$ is quasi-normalized both ways:

- without approximation, since

$$L_{\frac{18}{6}} + L_{\frac{8}{6}} + L_{\frac{0}{6}} + L_{\frac{4}{6}} + L_{\frac{2}{6}} + L_{\frac{4}{6}} = L_{\frac{18+8+0+4+2+4}{6}} = L_{\frac{36}{6}} = L_6 = L_{\max},$$

- and with approximations

$$L_{[\frac{18}{6}]} + L_{[\frac{8}{6}]} + L_{[\frac{0}{6}]} + L_{[\frac{4}{6}]} + L_{[\frac{2}{6}]} + L_{[\frac{4}{6}]} = L_3 + L_1 + L_0 + L_1 + L_0 + L_1 = L_6 = L_{\max}.$$

Compute $qm_{\text{Mix}}(\cdot)$ using the second similarity/dissimilarity weights given by (30) (which are equal in this case with Jaccard's distance similarity/dissimilarity weights). In this case, we get better results. Since from Table VIII and (28), one gets

$$\begin{aligned}
qm^{\eta}_{\text{Mix}}(A) &= \eta(A,A)qm_1(A)qm_2(A) + (1-\eta(A,A))qm_1(A)qm_2(A) + (1-\eta(A,A\cup B))qm_1(A)qm_2(A\cup B) + (1-\eta(A\cup B,A))qm_1(A\cup B)qm_2(A)\\
&= 0\cdot L_2L_3 + 1\cdot L_2L_3 + \tfrac{1}{2}\cdot L_2L_0 + \tfrac{1}{2}\cdot L_4L_3\\
&= L_0 + L_{\frac{2\cdot 3}{6}} + L_{\frac{2\cdot 0}{6\cdot 2}} + L_{\frac{4\cdot 3}{6\cdot 2}} = L_{0+\frac{6}{6}+0+\frac{6}{6}} = L_{\frac{12}{6}},\\
qm^{\eta}_{\text{Mix}}(B) &= (1-\eta(B,A\cup B))qm_1(B)qm_2(A\cup B)\\
&= \tfrac{1}{2}\cdot L_2L_4 = L_{\frac{4\cdot 2}{6\cdot 2}} = L_{\frac{4}{6}},\\
qm^{\eta}_{\text{Mix}}(A\cup B) &= \eta(A\cup B, A\cup B)qm_1(A\cup B)qm_2(A\cup B) + \eta(A, A\cup B)qm_1(A)qm_2(A\cup B) + \eta(A\cup B, A)qm_1(A\cup B)qm_2(A)\\
&\quad + \eta(B, A\cup B)qm_1(B)qm_2(A\cup B) + \eta(A\cup B, B)qm_1(A\cup B)qm_2(B) + \eta(A,B)qm_1(A)qm_2(B)\\
&\quad + \eta(B,A)qm_1(B)qm_2(A) + (1-\eta(A\cup B, A\cup B))qm_1(A\cup B)\cdot qm_2(A\cup B)\\
&= 0\cdot L_4L_0 + \tfrac{1}{2}\cdot L_2L_0 + \tfrac{1}{2}\cdot L_4L_3\,\tfrac{1}{2}\cdot L_0L_0 + \tfrac{1}{2}\cdot L_4L_2 + 1\cdot L_2L_2 + 1\cdot L_0L_3 + 1\cdot L_4L_0\\
&= L_0 + L_0 + L_{\frac{4\cdot 3}{6\cdot 2}} + L_0 + L_{\frac{4\cdot 2}{6\cdot 2}} + L_{\frac{2\cdot 2}{6}} + L_0\\
&= L_{\frac{6+4+4}{6}} = L_{\frac{14}{6}},
\end{aligned}$$

$$\begin{aligned}
qm^{\eta}_{\text{Mix}}(A\cup C) &= \eta(A,C)qm_1(A)qm_2(C) + (1-\eta(C,A))qm_1(C)qm_2(A)\\
&= 1\cdot L_2L_1 + 1\cdot L_0L_3 = L_{\frac{2\cdot 1}{6}} + L_{\frac{0\cdot 3}{6}} = L_{\frac{2}{6}},\\
qm^{\eta}_{\text{Mix}}(A\cup B\cup C) &= \eta(C, A\cup B)qm_1(C)qm_2(A\cup B) + \eta(A\cup B, C)qm_1(A\cup B)qm_2(C)\\
&= 1\cdot L_0L_0 + 1\cdot L_4L_1 = L_{\frac{4}{6}}.
\end{aligned}$$

Similarly, $qm^{\eta}_{\text{Mix}}(\cdot)$ is quasi-normalized both ways.

- **Discounted Qualitative PCR (36)**

We show how to apply the Discounted Qualitative PCR rule (36) in this example with the fixed discounting factor $\alpha = 0.6$, hence $1-\alpha = 0.4$. First, apply the qualitative conjunctive rule.

TABLE XVII
Qualitative Inputs and Conjunctive Rule

	A	B	C	$A\cup B$
$qm_1(\cdot)$	L_2	L_0	L_0	L_4
$qm_2(\cdot)$	L_3	L_2	L_1	L_0
$qm_{\wedge}(\cdot)$	$L_{\frac{18}{6}}$	$L_{\frac{8}{6}}$	L_0	L_0

Indeed, one has

$$\begin{aligned}
qm_{\wedge}(A) &= L_2L_3 + L_2L_0 + L_3L_4 = L_{\frac{2\cdot 3}{6}+0+\frac{3\cdot 4}{6}} = L_{\frac{18}{6}},\\
qm_{\wedge}(B) &= L_0L_2 + L_0L_0 + L_2L_4 = L_{0+0+\frac{2\cdot 4}{6}} = L_{\frac{8}{6}},\\
qm_{\wedge}(C) &= L_0L_1 = L_0,\\
qm_{\wedge}(A\cup B) &= L_4L_0 = L_0.
\end{aligned}$$

Applying the proportional conflict redistribution according to PCR, one has

$$\frac{x_{1,A}}{L_2} = \frac{y_{1,B}}{L_2} = \frac{L_2L_2}{L_2+L_2} = \frac{L_{\frac{4}{6}}}{L_4} = L_{(\frac{4}{6}\div 4)\cdot 6} = L_1.$$

Therefore,

$$\begin{aligned}
x_{1,A} &= L_2L_1 = L_{\frac{2}{6}},\\
y_{1,B} &= L_2L_1 = L_{\frac{2}{6}},\\
\frac{x_{2,A}}{L_2} &= \frac{z_{1,C}}{L_1} = \frac{L_2L_1}{L_2+L_1} = \frac{L_{\frac{2}{6}}}{L_3} = L_{(\frac{2}{6}\div 3)\cdot 6} = L_{\frac{4}{6}},\\
x_{2,A} &= L_2L_{\frac{4}{6}} = L_{\frac{4/3}{6}},\\
z_{1,C} &= L_1L_{\frac{4}{6}} = L_{\frac{2/3}{6}},\\
\frac{z_{2,C}}{L_1} &= \frac{w_{1,A\cup B}}{L_4} = \frac{L_1L_4}{L_1+L_4} = \frac{L_{\frac{4}{6}}}{L_5} = L_{(\frac{4}{6}\div 5)\cdot 6} = L_{\frac{4}{5}},
\end{aligned}$$

and

$$z_{2,C} = L_1 L_{\frac{4}{5}} = L_{\frac{1\cdot 4/5}{6}} = L_{\frac{0.8}{6}},$$
$$w_{1,A\cup B} = L_4 L_{\frac{4}{5}} = L_{\frac{4\cdot 4/5}{6}} = L_{\frac{3.2}{6}}.$$

Summing, we get

$$\begin{aligned}
qm_{\text{DPCR}}(A) &= L_{\frac{1}{6}} + 0.6 \cdot \left(L_{\frac{2}{6}} + L_{\frac{4/3}{6}}\right) \\
&= L_{\frac{1}{6}} + 0.6 \cdot L_{\frac{10/3}{6}} = L_{\frac{1}{6}} + L_{\frac{2}{6}} = L_{\frac{2}{6}}, \\
qm_{\text{DPCR}}(B) &= L_{\frac{8}{6}} + 0.6 \cdot (L_{\frac{2}{6}}) = L_{\frac{8}{6}} + L_{\frac{1.2}{6}} = L_{\frac{9.2}{6}}, \\
qm_{\text{DPCR}}(C) &= L_0 + 0.6 \cdot \left(L_{\frac{4/3}{6}} + L_{\frac{0.8}{6}}\right) \\
&= L_0 + L_{\frac{0.62/3+0.8}{6}} = L_{\frac{0.88}{6}}, \\
qm_{\text{DPCR}}(A\cup B) &= L_0 + 0.6 \cdot (L_{\frac{3.2}{6}} + 0.4 \cdot (L_2 L_2 + L_3 L_2) \\
&= L_{\frac{1.92}{6}} + 0.4 \cdot L_{\frac{2.2}{6}} = L_{\frac{1.92}{6}+\frac{1.60}{6}} = L_{\frac{3.52}{6}}, \\
qm_{\text{DPCR}}(A\cup C) &= 0.4 \cdot (L_2 L_1 + L_3 L_0) \\
&= 0.4 \cdot (L_{\frac{2.1}{6}} + L_0) = L_{\frac{0.8}{6}}, \\
qm_{\text{DPCR}}(A\cup B\cup C) &= 0.4 \cdot (L_0 L_0 + L_1 L_4) \\
&= 0.4 \cdot \left(L_0 + L_{\frac{1.4}{6}}\right) = L_{\frac{1.6}{6}}.
\end{aligned}$$

We can check that $qm_{\text{DPCR}}(\cdot)$ is quasi-normalized both ways.

- **Mixed Discounted Qualitative PCR (44)**

In this example, we still set the discounting factor to $\alpha = 0.6$.

1) Using the first kind of similarity/dissimilarity weights (see Table VI), one obtains

$$\begin{aligned}
qm^{\delta}_{\text{MDPCR}}(A) &= \delta_1(A,A) qm_1(A) qm_2(A) \\
&\quad + \delta_2(A,A) qm_1(A) qm_2(A) \\
&\quad + \delta_2(A, A\cup B) qm_1(A) qm_2(A\cup B) \\
&\quad + \delta_2(A\cup B, A) qm_1(A\cup B) qm_2(A) \\
&\quad + \alpha \cdot \left(L_{\frac{2}{6}} + L_{\frac{4/3}{6}}\right) \\
&= 0 \cdot L_2 L_3 + 1 \cdot L_2 L_3 + 1 \cdot L_2 L_0 \\
&\quad + 1 \cdot L_4 L_3 + 0.6 \cdot \left(L_{\frac{2}{6}} + L_{\frac{4/3}{6}}\right) \\
&= L_{\frac{1}{6}} + 0.6 \cdot L_{\frac{10/3}{6}} = L_{\frac{1}{6}} + L_{\frac{2}{6}} = L_{\frac{2}{6}}.
\end{aligned}$$

The term $L_{\frac{4/3}{6}}$ in the sum above comes from the previous Discounted Qualitative PCR example.
One gets the same result as in the previous example (Discounted Qualitative PCR).

2) Using the second kind of similarity/dissimilarity weights (see Table VIII), one obtains:

$$\begin{aligned}
qm^{\eta}_{\text{MDPCR}}(A) &= \eta(A,A) qm_1(A) qm_2(A) \\
&\quad + (1-\eta(A,A)) qm_1(A) qm_2(A) \\
&\quad + (1-\eta(A, A\cup B)) qm_1(A) qm_2(A\cup B) \\
&\quad + (1-\eta(A\cup B, A)) qm_1(A\cup B) qm_2(A) \\
&\quad + \alpha \cdot \left(L_{\frac{2}{6}} + L_{\frac{4/3}{6}}\right) \\
&= 0 \cdot L_2 L_3 + 1 \cdot L_2 L_3 + \tfrac{1}{2} \cdot L_2 L_0 \\
&\quad + \tfrac{1}{2} \cdot L_4 L_3 + 0.6 \cdot \left(L_{\frac{2}{6}} + L_{\frac{4/3}{6}}\right) \\
&= L_{\frac{1.2}{6}} + L_{\frac{2}{6}} = L_{\frac{1.4}{6}}.
\end{aligned}$$

Similarly

$$\begin{aligned}
qm^{\eta}_{\text{MDPCR}}(B) &= 0 \cdot L_0 L_2 + 1 \cdot L_0 L_2 + \tfrac{1}{2} \cdot L_0 L_0 \\
&\quad + \tfrac{1}{2} \cdot L_4 L_2 + 0.6 \cdot L_{\frac{2}{6}} \\
&= \tfrac{1}{2} \cdot L_4 L_2 + L_{\frac{1.2}{6}} = L_{\frac{4.2}{6.2}} + L_{\frac{1.2}{6}} \\
&= L_{\frac{4}{6}} + L_{\frac{1.2}{6}} = L_{\frac{5.2}{6}}, \\
qm^{\eta}_{\text{MDPCR}}(C) &= 0 \cdot L_0 L_1 + 1 \cdot L_0 L_1 + 0.6 \cdot \left(L_{\frac{2/3}{6}} + L_{\frac{0.8}{6}}\right) \\
&= L_0 + L_0 + L_{\frac{0.88}{6}} = L_{\frac{0.88}{6}}.
\end{aligned}$$

The term $L_{\frac{0.8}{6}}$ in the sum above comes from the previous Discounted Qualitative PCR example. We get

$$\begin{aligned}
qm^{\eta}_{\text{MDPCR}}(A\cup B) &= \eta(A\cup B, A\cup B) qm_1(A\cup B) \\
&\quad \cdot qm_2(A\cup B) \\
&\quad + \eta(A, A\cup B) qm_1(A) qm_2(A\cup B) \\
&\quad + \eta(A\cup B, A) qm_1(A\cup B) qm_2(A) \\
&\quad + \eta(B, A\cup B) qm_1(B) qm_2(A\cup B) \\
&\quad + \eta(A\cup B, B) qm_1(A\cup B) qm_2(B) \\
&\quad + (1-\eta(A\cup B, A\cup B)) qm_1(A\cup B) \\
&\quad \cdot qm_2(A\cup B) \\
&\quad + \alpha \cdot L_{\frac{3.2}{6}} + (1-\alpha) qm_1(A) qm_2(B) \\
&\quad + (1-\alpha) qm_1(B) qm_2(A), \\
&= 0 \cdot L_4 L_0 + \tfrac{1}{2} \cdot L_0 L_1 + \tfrac{1}{2} \cdot L_4 L_3 \tfrac{1}{2} \cdot L_0 L_0 \\
&\quad + \tfrac{1}{2} \cdot L_4 L_2 + 1 \cdot L_4 L_0 \\
&\quad + 0.6 \cdot L_{\frac{3.2}{6}} + 0.4 \cdot L_2 L_2 + 0.4 \cdot L_0 L_3 \\
&= L_{\frac{4.3}{6.2}} + L_{\frac{4.2}{6.2}} + L_{\frac{1.92}{6}} + L_{\frac{1.60}{6}} \\
&= L_{\frac{6}{6}} + L_{\frac{4}{6}} + L_{\frac{1.92}{6}} + L_{\frac{1.60}{6}} = L_{\frac{13.52}{6}},
\end{aligned}$$

$$qm^{\eta}_{\text{MDPCR}}(A\cup C) = (1-\alpha)qm_1(A)qm_2(C) + (1-\alpha)qm_1(C)qm_2(A) = 0.4\cdot L_2L_1 + 0.4\cdot L_3L_0 = L_{\frac{0.8}{6}},$$

$$qm^{\eta}_{\text{MDPCR}}(A\cup B\cup C) = (1-\alpha)qm_1(C)qm_2(A\cup B) + (1-\alpha)qm_1(A\cup B)qm_2(C) = 0.4\cdot L_0L_0 + 0.4\cdot L_4L_1 = L_{\frac{1.6}{6}}.$$

$qm^{\eta}_{\text{MDPCR}}(\cdot)$ is quasi-normalized without approximations, but it is not with approximations.

EXAMPLE 13 Fusion of three sources

Consider a set of labels $L = \{L_{\min} = L_0, L_1, L_2, L_3, L_4, L_5, L_{\max} = L_6\}$, and a frame of discernment $\Theta = \{A, B, C\}$ in Shafer's model (*i.e.* all intersections empty). Let's take the three following qualitative sources of evidence described in Table XVIII.

TABLE XVIII
Qualitative Inputs for Example 13

	A	B	$B\cup C$	$A\cup B\cup C$
$qm_1(\cdot)$	L_2	L_0	L_0	L_4
$qm_2(\cdot)$	L_0	L_3	L_0	L_3
$qm_3(\cdot)$	L_0	L_0	L_5	L_1

- **Qualitative conjunctive rule**

If one applies the Qualitative Conjunctive Rule (QCR), one gets

$$qm_{\wedge}(A) = qm_1(A)qm_2(A\cup B\cup C\cup D)qm_3(A\cup B\cup C\cup D) = L_2L_3L_1 = L_{\frac{2\cdot 3}{6}}L_1 = L_{\frac{2\cdot 3\cdot 1}{6\cdot 6}} = L_{\frac{1}{6}}.$$

Similarly,

$$qm_{\wedge}(B) = L_4L_3L_1 + L_4L_3L_5 = L_{\frac{4\cdot 3\cdot 1}{6\cdot 6}} + L_{\frac{4\cdot 3\cdot 5}{6\cdot 6}} = L_{\frac{2}{6}} + L_{\frac{10}{6}} = L_{\frac{12}{6}},$$

$$qm_{\wedge}(B\cup C) = L_4L_3L_5 = L_{\frac{4\cdot 3\cdot 5}{6\cdot 6}} = L_{\frac{10}{6}},$$

$$qm_{\wedge}(A\cup B\cup C\cup D) = L_4L_3L_1 = L_{\frac{4\cdot 3\cdot 1}{6\cdot 6}} = L_{\frac{2}{6}}.$$

The total conflict is

$$qm_{\wedge}(\emptyset) = qm_1(A)qm_2(B)qm_3(B\cup C) + qm_1(A)qm_2(B)qm_3(A\cup B\cup C\cup D) + qm_1(A)qm_2(A\cup B\cup C\cup D)qm_3(B\cup C) = L_2L_3L_5 + L_2L_3L_1 + L_2L_3L_5 = L_{\frac{2\cdot 3\cdot 5}{6\cdot 6}} + L_{\frac{2\cdot 3\cdot 1}{6\cdot 6}} + L_{\frac{2\cdot 3\cdot 5}{6\cdot 6}} = L_{\frac{5}{6}} + L_{\frac{1}{6}} + L_{\frac{5}{6}} = L_{\frac{11}{6}}.$$

- **Qualitative PCR**

Applying the proportional conflict redistribution for the first partial conflict $qm_1(A)qm_2(B)qm_3(B\cup C)$, one gets

$$\frac{x_{1,A}}{L_2} = \frac{y_{1,B}}{L_3} = \frac{z_{1,B\cup C}}{L_5} = \frac{L_2L_3L_5}{L_2+L_3+L_5} = \frac{L_{\frac{5}{6}}}{L_{10}} = L_{(\frac{5}{6}\div 10)\cdot 6} = L_{\frac{3}{6}}.$$

Therefore,

$$x_{1,A} = L_2L_{\frac{3}{6}} = L_{\frac{2\cdot 3}{6\cdot 6}} = L_{\frac{1}{6}},$$

$$y_{1,B} = L_3L_{\frac{3}{6}} = L_{\frac{3\cdot 3}{6\cdot 6}} = L_{\frac{1.5}{6}},$$

$$z_{1,B\cup C} = L_5L_{\frac{3}{6}} = L_{\frac{5\cdot 3}{6\cdot 6}} = L_{\frac{2.5}{6}}.$$

Applying the proportional conflict redistribution for the second partial conflict $qm_1(A)qm_2(B)qm_3(A\cup B\cup C\cup D)$, one gets

$$\frac{x_{2,A}}{L_2} = \frac{y_{2,B}}{L_3} = \frac{w_{1,A\cup B\cup C\cup D}}{L_1} = \frac{L_2L_3L_1}{L_2+L_3+L_1} = \frac{L_{\frac{1}{6}}}{L_6} = L_{(\frac{1}{6}\div 6)\cdot 6} = L_{\frac{1}{6}}.$$

Therefore,

$$x_{2,A} = L_2L_{\frac{1}{6}} = L_{\frac{2\cdot 1}{6\cdot 6}} = L_{\frac{1/3}{6}},$$

$$y_{2,B} = L_3L_{\frac{1}{6}} = L_{\frac{3\cdot 1}{6\cdot 6}} = L_{\frac{1/2}{6}} = L_{\frac{0.5}{6}},$$

$$w_{1,A\cup B\cup C\cup D} = L_1L_{\frac{1}{6}} = L_{\frac{1\cdot 1}{6\cdot 6}} = L_{\frac{1/6}{6}}.$$

Applying the proportional conflict redistribution for the third partial conflict $qm_1(A)qm_2(A\cup B\cup C\cup D)\cdot qm_3(B\cup C)$, one gets

$$\frac{x_{3,A}}{L_2} = \frac{w_{2,A\cup B\cup C\cup D}}{L_3} = \frac{z_{2,B\cup C}}{L_5} = \frac{L_2L_3L_5}{L_2+L_3+L_5} = L_{\frac{3}{6}},$$

so,

$$x_{3,A} = L_2L_{\frac{3}{6}} = L_{\frac{1}{6}},$$

$$w_{2,A\cup B\cup C\cup D} = L_3L_{\frac{3}{6}} = L_{\frac{1.5}{6}},$$

$$z_{2,B\cup C} = L_5L_{\frac{3}{6}} = L_{\frac{2.5}{6}}.$$

Summing, we get

$$qm_{\text{PCR}}(A) = L_{\frac{1}{6}} + L_{\frac{1}{6}} + L_{\frac{1/3}{6}} + L_{\frac{1}{6}} = L_{\frac{10/3}{6}},$$

$$qm_{\text{PCR}}(B) = L_{\frac{12}{6}} + L_{\frac{1.5}{6}} + L_{\frac{0.5}{6}} = L_{\frac{14}{6}},$$

$$qm_{\text{PCR}}(B\cup C) = L_{\frac{10}{6}} + L_{\frac{2.5}{6}} + L_{\frac{2.5}{6}} = L_{\frac{15}{6}},$$

$$qm_{\text{PCR}}(A\cup B\cup C\cup D) = L_{\frac{2}{6}} + L_{\frac{1/6}{6}} + L_{\frac{1.5}{6}} = L_{\frac{22/6}{6}}.$$

We can check that $qm_{\text{PCR}}(\cdot)$ is quasi-normalized without approximations (*i.e.* when working within the refined set of linguistic labels by keeping fractional indexes), but it is not quasi-normalized when using approximations of fractional indexes if we want to work back within the original set of linguistic labels $L = \{L_{\min} = L_0, L_1, L_2, L_3, L_4, L_5, L_{\max} = L_6\}$.

• **Discounted Qualitative PCR (37)**

Let's consider the discounting factor $\alpha = 0.6$. Consider the previous example and discount it according to (37) applied in the qualitative domain. One obtains:

$$\begin{aligned}
qm_{\mathrm{DPCR}}(A) &= L_{\frac{1}{6}} + 0.6 \cdot \left(L_{\frac{1}{6}} + L_{\frac{1/3}{6}} + L_{\frac{1}{6}} \right) \\
&= L_{\frac{1}{6}} + 0.6 \cdot L_{\frac{7/3}{6}} = L_{\frac{2.4}{6}}, \\
qm_{\mathrm{DPCR}}(B) &= L_{\frac{12}{6}} + 0.6 \cdot \left(L_{\frac{1.5}{6}} + L_{\frac{0.5}{6}} \right) \\
&= L_{\frac{1}{6}} + 0.6 \cdot L_{\frac{2}{6}} = L_{\frac{13.2}{6}}, \\
qm_{\mathrm{DPCR}}(B \cup C) &= L_{\frac{10}{6}} + 0.6 \cdot \left(L_{\frac{2.5}{6}} + L_{\frac{2.5}{6}} \right) \\
&= L_{\frac{10}{6}} + 0.6 \cdot L_{\frac{5}{6}} = L_{\frac{13}{6}}, \\
qm_{\mathrm{DPCR}}(A \cup B \cup C) &= (1-\alpha) qm_1(A) qm_2(B) qm_2(B \cup C) \\
&= 0.4 \cdot L_2 L_3 L_5 = 0.4 \cdot L_{\frac{5}{6}} = L_{\frac{2}{6}}, \\
qm_{\mathrm{DPCR}}(A \cup B \cup C \cup D) &= L_{\frac{2}{6}} + 0.6 \cdot \left(L_{\frac{1/6}{6}} + L_{\frac{1.5}{6}} \right) \\
&\quad + 0.4 \cdot (L_2 L_3 L_1 + L_2 L_3 L_5) \\
&= L_{\frac{2}{6}} + 0.6 \cdot L_{\frac{5/3}{6}} + 0.4 \cdot \left(L_{\frac{1}{6}} + L_{\frac{5}{6}} \right) \\
&= L_{\frac{2}{6}} + L_{\frac{1}{6}} + 0.4 \cdot L_{\frac{6}{6}} \\
&= L_{\frac{3}{6}} + L_{\frac{2.4}{6}} = L_{\frac{5.4}{6}}.
\end{aligned}$$

$qm_{\mathrm{DPCR}}(\cdot)$ is quasi-normalized without approximations, but it is not with approximations.

7. CONCLUSIONS

With the recent development of qualitative methods for reasoning under uncertainty developed in Artificial Intelligence, more and more experts and scholars have expressed great interest on qualitative information fusion, especially those working in the development of modern multi-source systems for defense, robot navigation, mapping, localization and path planning and so on. In this paper, we propose some solutions to handle the conflict and to weigh the imprecision of the responses of the experts, from the classical combination rules for qualitative and quantitative beliefs. Hence, we have presented a mixed rule given by a weighted sum of the conjunctive and disjunctive rules. The weights are defined from a measure of non-specifity calculated by the cardinality of the responses of the experts. This rule transfers the partial conflict on partial ignorance. Again, the proportional conflict distribution rule redistributes the partial conflict on the element implied in this conflict. We propose an extension of this rule by a discounting procedure, thereby, a part of the partial conflict is also redistributed on the partial ignorance. We have introduced a measure of conflict between pair of experts and another measure of non-conflict between pair of experts, as to quantify this part. In order to take heed of the non-specifity and to redistributed the partial conflict, we propose a fused rule of these two new rules. This rule is created in such way that we retain a kind of continuity of the mass on the partial ignorance, between both cases with and without partial conflict. Illustrating examples have been presented in detail to explain how the new rules work for quantitative and qualitative beliefs. The study of these new rules shows that the classical combination rules in the belief functions theory cannot take precisely into account the non-specifity of the experts and the partial conflict of the experts. This is specially important for qualitative belief.

REFERENCES

[1] A. Appriou
Approche générique de la gestion de l'incertain dans les processus de fusion multisenseur.
Traitement du Signal, **22**, 4 (2005), 307–319.

[2] A. P. Dempster
Upper and Lower probabilities induced by a multivalued mapping.
Annals of Mathematical Statistics, **83** (1967), 325–339.

[3] A. P. Dempster
A generalization of Bayesian inference.
Journal of the Royal Statistical Society, Serie B, **30** (1968), 205–245.

[4] T. Denœux
The cautious rule of combination for belief functions and some extensions.
International Conference on Information Fusion, Florence, Italy, July 10–13, 2006.

[5] D. Dubois and H. Prade
Representation and combination of uncertainty with belief functions and possibility measures.
Computational Intelligence, **4** (1988), 244–264.

[6] D. Dubois and H. Prade
Possibility Theory: An Approach to Computerized Processing of Uncertainty.
New York: Plenum Press, 1988.

[7] M. C. Florea, J. Dezert, P. Valin, F. Smarandache and A. L. Jousselme
Adaptative combination rule and proportional conflict redistribution rule for information fusion.
COGnitive systems with Interactive Sensors, Paris, France, Mar. 2006.

[8] T. Inagaki
Independence between safety-control policy and multiple-sensors schemes via Dempster-Shafer theory.
IEEE Transaction on Reliability, **40**, 2 (1991), 182–188.

[9] A. Josang, M. Daniel and P. Vannoorenberghe
Strategies for combining conflicting dogmatic belief.
International Conference on Information Fusion, Cairns, Australia, July 7–10, 2003.

[10] A.-L. Jousselme, D. Grenier and E. Bossé
A new distance between two bodies of evidence.
Information Fusion, **2** (2001), 91–101.

[11] L. Lam and C. Y. Suenot
Application of majority voting to pattern recognition: An analysis of its behavior and performance.
IEEE Transactions on Systems, Man Cybernetics—Part A: Systems and Humans, **27**, 5 (Sept. 1997), 553–568.

[12] E. Lefevre, O. Colot and P. Vannoorenberghe
Belief function combination and conflict management.
Information Fusion, **3** (2002), 149–162.

[13] X. Li, X. Huang, J. Dezert and F. Smarandache
Enrichment of qualitative belief for reasoning under uncertainty.
International Conference on Information Fusion, Québec City, Canada, July 9–12, 2007.

[14] A. Martin and C. Osswald
A new generalization of the proportional conflict redistribution rule stable in terms of decision.
Applications and Advances of DSmT for Information Fusion, Book 2, American Research Press, Rehoboth, F. Smarandache and J. Dezert (Eds.), 2006, 69–88.

[15] A. Martin and C. Osswald
Toward a combination rule to deal with partial conflict and specificity in belief functions theory.
International Conference on Information Fusion, Québec, Canada, July 9–12, 2007.

[16] C. Osswald and A. Martin
Understanding the large family of Dempster-Shafer theory's fusion operators—a decision-based measure.
International Conference on Information Fusion, Florence, Italy, July 10–13, 2006.

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

[18] F. Smarandache and J. Dezert (Eds.)
Applications and Advances of DSmT for Information Fusion (Collected works).
American Research Press, Rehoboth, 2004. http://www.gallup.unm.edu/∼smarandache/DSmT-book1.pdf.

[19] F. Smarandache and J. Dezert (Eds.)
Applications and Advances of DSmT for Information Fusion (Collected works, vol. 2.)
American Research Press, Rehoboth, 2006. http://www.gallup.unm.edu/∼smarandache/DSmT-book2.pdf.

[20] F. Smarandache and J. Dezert (Eds.)
Applications and Advantages of DSmT for Information Fusion (Collected works, vol. 3).
American Research Press, Rehoboth, 2009, http://www.gallup.unm.edu/∼smarandache/DSmT-book3.pdf.

[21] F. Smarandache and J. Dezert
Information fusion based on new proportional conflict redistribution rules.
International Conference on Information Fusion, Philadelphia, PA, July 25–29, 2005.

[22] F. Smarandache and J. Dezert
Proportional conflict redistribution rules for information fusion.
Applications and Advances of DSmT for Information Fusion, Book 2, American Research Press, Rehoboth, F. Smarandache and J. Dezert (Eds.), 2006, 3–68.

[23] F. Smarandache
An in-depth look at quantitative information fusion rules.
In *Applications and Advances of DSmT for Information Fusion*, Book 2, American Research Press, Rehoboth, F. Smarandache and J. Dezert (Eds.), 2006, 205–235 (originally presented at NASA Langley Research Center, Hampton, VA, on Nov. 5, 2005).

[24] F. Smarandache and J. Dezert
Qualitative belief conditioning rules (QBCR).
International Conference on Information Fusion, Québec City, Canada, July 9–12, 2007.

[25] Ph. Smets
Constructing the pignistic probability function in a context of uncertainty.
Uncertainty in Artificial Intelligence, **5** (1990), 29–39.

[26] Ph. Smets
The combination of evidence in the transferable belief model.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **12**, 5 (1990), 447–458.

[27] Ph. Smets
Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem.
International Journal of Approximate Reasoning, **9** (1993), 1–35.

[28] Ph. Smets and R. Kennes
The transferable belief model.
Artificial Intelligent, **66** (1994), 191–234.

[29] Ph. Smets
The α-junctions: The commutative and associative non interactive combination operators applicable to belief functions.
Qualitative and Quantitative Practical Reasoning, Springer Verlag, D. Gabbay and R. Kruse and A. Nonnengart and H. J. Ohlbacg, Berlin, 1997, 131–153.

[30] K. Sentz and S. Ferson
Combination of Evidence in Dempster-Shafer Theory.
SANDIA Tech. Report, SAND2002-0835, Apr. 2002. 96 pages.

[31] L. Xu, A. Krzyzak and C. Y. Suen
Methods of combining multiple classifiers and their application to handwriting recognition.
IEEE Transactions on Systems, Man Cybernetics, **22**, 3 (May 1992), 418–435.

[32] R. R. Yager
Hedging in the combination of evidence.
Journal of Information and Optimization Science, **4**, 1 (1983), 73–81.

[33] R. R. Yager
On the relationships of methods of aggregation of evidence in expert systems.
Cybernetics and Systems, **16** (1985), 1–21.

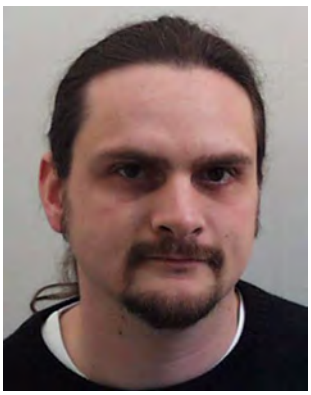
[34] R. R. Yager
On the Dempster-Shafer framework and new combination rules.
Informations Sciences, **41** (1987), 93–137.

[35] L. Zadeh
Fuzzy sets as a basis for a theory of possibility.
Fuzzy Sets and Systems, **1**, 3 (1978), 3–28.

[36] L. Zadeh
On the validity of Dempster's rule of combination.
Memo M 79/24, Univ. of California, Berkeley, 1979.

Arnaud Martin was born in Bastia, France in 1974. He received the Ph.D. degree in signal processing (2001), and the M.S. degree in probability (1998) from the University of Rennes, France.

He worked on speech recognition for three years (1998–2001) at France Telecom R&D, Lannion, France. He worked in the Department of Statistics and Data Mining (STID) of the IUT of Vannes, France, as temporary assistant professor (ATER) for two years (2001–2003). In 2003, he joined the Laboratory E^3I^2: EA3876 at the ENSIETA, Brest, France, as a teacher and researcher. Dr. Arnaud Martin teaches mathematics, data fusion, data mining, signal processing, and computer sciences. Today, his research interests are mainly related to the belief functions for pattern recognition of real data and include data fusion, data mining, signal processing, especially for sonar and radar data.



Christophe Osswald was born in Orléans, France, in 1975. He received the Ph.D. degree in mathematics and computer science (2003) from the EHESS—École des Hautes Études en Sciences Sociales—after three years of research on classification, similarity analysis, and hypergraphs at the Laboratory IASC of the ENST Bretagne. He is an engineer graduated from the École Polytechnique (1997) and the ENST Bretagne (1999), and received the M.S. degree in mathematics and computer science applied to social sciences.

He is a teacher and researcher at the ENSIETA in the Laboratory E^3I^2: EA3876, Brest, France. His research interests are related to data fusion, mostly through belief functions, and also to classification, graphs and hypergraphs theory, algorithmic complexity, and decision support.

Jean Dezert was born in l'Hay les Roses, France, on August 25, 1962. He received the electrical engineering degree from the Ecole Française de Radioélectricité Electronique and Informatique (EFREI), Paris, in 1985, the D.E.A. degree in 1986 from the University Paris VII (Jussieu), and his Ph.D. from the University Paris XI, Orsay, in 1990, all in automatic control and signal processing.

During 1986–1990 he was with the Systems Department at the French Aerospace Research Lab (ONERA), Châtillon, France, and did research in multisensor multitarget tracking (MS-MTT). During 1991–1992, he visited the Department of Electrical and Systems Engineering, University of Connecticut, Storrs, as a European Space Agency (ESA) Postdoctoral Research Fellow. During 1992–1993 he was teaching assistant in electrical engineering at the University of Orléans, France. Since 1993, he is a senior research scientist in the Information Modelling and Processing Department (DTIM) at ONERA. His current research interests include autonomous navigation, estimation theory, stochastic systems theory and its applications to MS-MTT, information fusion, plausible reasoning and non-standard Logics. Dr. Jean Dezert is developing since 2001 with Professor Smarandache a new theory of plausible and paradoxical reasoning for information fusion (DSmT) and has edited two textbooks (collected works) devoted to this new emerging research field published by American Research Press, Rehoboth in 2004 and 2006 respectively and a third volume is in preparation (see http://www.gallup.unm.edu/∼smarandache/DSmT.htm). He owns one international patent in the autonomous navigation field and has published several papers tracking and fusion in international conferences and journals. He was member of IEEE and of Eta Kappa Nu, serves as reviewer for different International Journals, taught courses on MS-MTT and Data Fusion at the French ENSTA Engineering School, collaborates for the development of the International Society of Information Fusion (ISIF) since 1998, and has served as Local Arrangements Organizer for Fusion 2000 Conference in Paris. He has been involved in the Technical Program Committees of Fusion 2001–2007 International Conferences. Since 2001, he is a member of the board of the International Society of Information Fusion (http://www.isif.org) and serves in ISIF executive board. He served as executive vicepresident of ISIF in 2004. Dr. Dezert gave several invited seminars, lectures and tutorials on Data Fusion and Tracking during recent past years—the last recent was given at Fusion 2008 in Cologne, Germany on July 2008.

Florentin Smarandache is a Professor and Chair of Math & Sciences Department at the University of New Mexico. He published over 75 books and 150 articles and notes in mathematics, physics, philosophy, psychology, literature, and rebus. In mathematics, his research is in number theory, non-Euclidean geometry, synthetic geometry, algebraic structures, statistics, neutrosophic logic and set (generalizations of fuzzy logic and set respectively), and neutrosophic probability (generalization of classical and imprecise probability). Other contributions are to nuclear and particle physics, neutrosophy (a generalization of dialectics), law of sensations and stimuli, and especially to information fusion.

Hierarchical Higher Level Data Fusion using Fuzzy Hamming and Hypercube Clustering

KEDAR SAMBHOOS
RAKESH NAGI
MOISES SUDIT
TERRY RICKARD

The primary objective of this research is to show progression of Level 2/3 fusion of informational content to obtain a framework for hierarchical high level decision making process. The goal of this paper is to develop an algorithm to solve the impending problem of situation assessment. In [31], an inexact graph matching technique called Truncated Search Tree algorithm (*TruST*) has been developed. The inexact graph matching is used to identify meaningful patterns in volumous amounts of data. This heuristic is based on the popular branch-and-bound technique with constraints on breadth and depth. To reduce the dimensionality of the matches found, the results are grouped using a clustering algorithm. A novel Hypercube distance measure is used in clustering the matched subgraphs. This measure is then compared with a relatively new Fuzzy Hamming distance measure. To identify the important nodes and links in the data graph, the clustered subgraphs are then fused together and the neighborhood structure is explored.

Manuscript received December 18, 2007; released for publication August 26, 2008.

Refereeing of this contribution was handled by Dr. David Hall.

This research is supported by Office of Naval Research and Lockheed-Martin Corporation, Conceptual Spaces and Graph Matching For High Level Fusion (Internal Research and Development).

Authors' addresses: Kedar Sambhoos, CUBRC, Buffalo, New York, E-mail: (sambhoos@cubrc.org); Rakesh Nagi and Moises Sudit, Center for MultiSource Information Fusion and Department of Industrial and Systems Engineering, State University of New York at Buffalo, Buffalo, New York; Terry Rickard, Lockheed Martin Corporation, Larkspur, Colorado.

1557-6418/08/$17.00

1. INTRODUCTION

Currently, analysts receive enormous amount of information from multiple sensors and resources. In addition to knowledge bases and traditional databases, enormous volumes of unstructured information are now available and are presented to the analyst [1], [35], [26]. This presents a significant challenge since different pieces of information from several heterogeneous data sources must be fused in order to generate conclusions relevant to the task at hand. The analyst is then left to sift through the information to analyze and interpret their contents and infer (uncover) what he or she is seeking. This process can be tedious and time-consuming and in many cases impractical when important events are unfolding rapidly. An automated means for this extraction of relevant pieces of information and their unification into a useful product for the analyst is much needed.

In order to address the problems of data gathering, fusion and extraction, The Joint Directors of Laboratories (JDL) Data Fusion Sub-panel developed a five-level Data Fusion Model [36]. Level 1 on Object refinement seems to have received the most attention. Level one processing functions include: data alignment, association, tracking, and identification. Less mature are Level 2 processing [17], [30], situation assessment, which seeks a higher level of inference above level one processing, and Level 3 processing which performs threat/impact assessment. Threat assessment is an iterative process of fusing the combined activity and capability of enemy forces to infer their intentions and assess the threat that they pose. Level 1 is very often called as "low-level" processing, and the others as "high-level" processing.

Higher level fusion problems are generally more difficult than level 1 because they involve higher dimensionality corresponding to the relationships among entities identified at level 1. Higher level fusion also concerns modeling behavior of aggregate entities, through the understanding of their individual behaviors and relationships. Some commonly recognized relationships are spatio-temporal relationships, part/whole relationships, organizational relationships, various causal relationships, semantic relationships, similarity relationships, etc. To represent these relationships we use (ARGs) Attributed Relational Graphs [11]. ARGs are rich data structures to represent level 1 objects along with higher level relationships as outlined above. In these ARGs, nodes represent people, organizations, location, individuals or facilities. Edges represent relationships like communication, radio, electronic, or telephone. Attributes store the details of each node and edge, like a individual's name or an event's time of occurrence. When we represent data as ARGs the problem of graph matching is of most importance when retrieving relevant information.

In this research we enhance level 2 and 3 fusion capability through a new class of hierarchical models

and algorithms based on graph representations and techniques. In [31], a method for imperfect subgraph matching [6] between a Template Graph and Data Graph is developed. This method produces multiple graphs corresponding to the alternate optimal or near optimal matches defined by the maximum value of similarity criterion. Depending on the truncation parameters used and the dimension of the data graph a large number of matches may be found. A number of these matches differ from each other in minor aspects reflected by one (or a few) nodes/arcs. It is difficult for the analyst to browse through each match to ascertain the distinguishing characteristics that characterize the set of matches. To reduce dimensionality of the matches found and to provide a hierarchical aggregation of these matches, a distance measure is employed between the matching graphs which is then used to group them using clustering algorithms. We employ two distance measures: Hypercube graph distance and Fuzzy Hamming distance [19], and use a simple K-means algorithm for clustering. We are also interested in the difference between the performance of the two distance measures. Finally, once we have obtained a cluster of similar matches, we would like to aggregate the match information for higher level interpretation.

The remainder of the paper is organized as follows. Section 2 is a brief literature review on graph distances. Section 3 presents the TruST algorithm [31]. Section 4 presents the K-means clustering using the Hypercube distance, while Section 5 presents the K-means clustering using the Fuzzy Hamming distance. The numerical comparison of the two distance measures using silhouette index is presented in Section 6. Section 7 presents the aggregation of clusters while Section 8 explores the neighborhood structure of the matches and clusters. Finally the conclusions are presented in Section 9.

2. LITERATURE REVIEW

There is a need to understand the situation and infer the enemy intent in the battlefield. To accomplish this the analyst needs to analyze the data at hand, to get an idea of the environment. There are many computational techniques implemented in high level data fusion (level 2 and 3). To mention some of them: Knowledge based expert systems [3], [2], Graph based matching techniques [12] [18], Bayesian belief systems approach [13], [14], Fuzzy Logic approach [27] [37], Genetic algorithms approach [4], [5], Artificial neural systems approach [24], [10]. The main concentration of this paper is on using the large number of matches (results) from TruST algorithm and providing a concise report to the analyst. Here we target most of the paper on using the fuzzy measures to find distance between graphs and cluster the matches. Being this target the literature survey concentrates mainly on graph distance measures. More detailed literature survey on Graph matching algorithms is being provided in [31].

There are currently many approaches in finding the similarity based distance between graphs. One such approach, suggested by Bunke and Shearer [8], uses a maximum common subgraph isomorphism algorithm to identify the largest substructure common to a pair of graphs, with the size of this maximum common subgraph (MCS) being determined by some function of the numbers of common vertices and edges. The distance metric between graphs G_1 and G_2, based on maximum common subgraph is given as:

$$d(G_1,G_2) = 1 - \frac{m(G_1,G_2)}{\max(|G_1|,|G_2|)} \tag{1}$$

where $m(G_1,G_2) = |G_{12}|$ is the number of vertices of the maximal common subgraph (G_{12}) of G_1 and G_2.

This provides a natural way of calculating the degree of similarity between a pair of graphs but the NP-complete nature of the maximum common subgraph isomorphism problem rules out the large-scale use of MCS-based similarities. Fernández and Valiente [15] extended the MCS based distance metric by combining maximum common subgraph and minimum common supergraph. This distance metric does not depend on the edit operations and is formulated as:

$$d(G_1,G_2) = |\check{G}| - |\hat{G}| \tag{2}$$

where $\check{G}$ and $\hat{G}$ are respectively, the maximum common subgraph and minimum common supergraph of G_1 and G_2. Wallis *et al.* [38] extended the MCS by defining the problem size using the union of two graphs being measured rather than the larger of the two graphs being used currently. The distance metric is formulated as follows:

$$d(G_1,G_2) = 1 - \frac{m(G_1,G_2)}{(|G_1| + |G_2| - |G_{12}|)}. \tag{3}$$

Chartrand *et al.* [9] and Kubicka and Kubicki [25] have defined the distance between graphs of equal order and size in terms of edge rotation and deletion. The authors [25] have shown that the distance is a metric and its application in matching a template graph with a database of graphs. The algorithm is limited to planar graphs and has the problem of computational explosion. The MCS algorithm and its variants are NP-hard and can be computationally expensive for large scale graphs. To avoid these exorbitant calculations, fuzzy distance measures are employed. In these methods the graphs are represented as vectors or points in a Hypercube and the distance is calculated based on the fuzzy theoretic measures. Fuzzy operations are computationally less expensive compared to MCS algorithms. A few fuzzy distance measures are addressed subsequently in Section 4 and Section 5.

One such interesting fuzzy distance metric is suggested by Ionescu and Ralescu [19], [20], [21], [22] called the (FHD) Fuzzy Hamming Distance. FHD metric name is also used by Bookstein *et al.* [7] to describe a

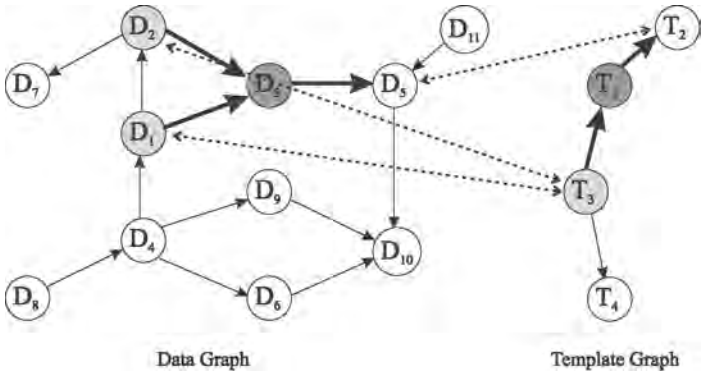


Fig. 1. Neighborhood scoring.

fuzzy distance measure based on edit distance cost. The metric used in this paper is the one devised by Ionescu and Ralescu. They have implemented this technique in banknote validator [22], image retrieval system [20] and image partitioning [19], [21]. The detail description of FHD is given in Section 5.

In this paper we use the inexact graph matching based heuristic [31], where through the control of truncation parameters we can control the state space. The use of truncated branch-and-bound as a heuristic gives better results [39] than most of the heuristics. This heuristic produces a large number of results and the analyst can be over whelmed from them, so the resulting matches are clustered. To cluster these results a novel Hypercube distance metric is being suggested and compared with a Fuzzy Hamming Distance measure. Section 3 gives details of the suggested heuristic, while Sections 4 and 5 respectively, give details on the Hypercube distance metric and Fuzzy Hamming distance metric.

3. TRUNCATED SEARCH TREE (TRUST) [31]

The TruST algorithm [31] expands on the idea of 1-Hop neighborhood distance. These 1-Hop neighborhoods consist of a root node and all other nodes of edge distance 1 away. For a pair of template node and data graph node, a linear assignment problem is solved over their neighborhood. The linear assignment problem takes the score matrix with individual elements as the average of the neighboring node-to-node score and connecting edge-to-edge score. So each 1-Hop neighborhood score will be a unique score depending on its corresponding neighborhood.

We define two parameters of the procedure to increase the flexibility of the solution and allow the algorithm to be configured for different domains of data.

1. α (**root weight score**): weights the value of the match between neighborhood root versus the value of the assignment of its neighbors. This will be discussed in more detail later.
2. t (**score threshold**): The algorithm will not return matches with a value below this threshold. The higher this threshold is set, the fewer 1-Hop neighborhood assignments would be determined, which in turn leads to improved performance (but potential loss of optimality).

Step 1 of the procedure is to compute a node-to-node score (C_{ij}) for each node in the template graph to each node in the data graph. For each node we then sort this list in descending order. Using the threshold value (t) input to the algorithm, we can prune the amount of assignments we must run for this template node by the equation $(t-1+\alpha)/\alpha$ (See proof in [31]). Root node scores which are below this value do not have the possibility of having an overall score above the threshold even when there is a perfect neighborhood assignment score.

Step 2 of the procedure is to compute the scores for the 1-Hop neighbors of each root node pair. This returns the optimal assignment of neighbors of the root node in the template graph to the neighbors of the root node in the data graph. Then the neighborhood score (shown in Fig. 1) between template graph node i and data graph node j is given by $(\alpha \times C_{ij} + (1-\alpha) \times W_{ij})$, where C_{ij} is the score of the root node pair and W_{ij} is the score of the neighborhood assignment, which is given by (4). α characterizes the amount of weightage to be given to the neighbors. α varies between $[0,1]$, and its value has inverse proportion with the neighborhood assignment.

$$W_{ij} = \frac{\text{Sum of neighborhood assignment scores}}{\text{Number of neighbors to root template node}}. \tag{4}$$

1-hop neighborhood score takes care of "node-to-node" assignment as well as "edge-to-edge" assignment. The score of neighborhood assignment is solved using a linear assignment problem between the root nodes with the adjacent nodes and edges forming the solution matrix. Using the results of the above two steps, a truncated search tree algorithm for matching is formulated.

A typical example of truncated search tree algorithm is shown in Fig. 2. The search tree is developed dynamically during the search and initially consists of only the root. At each iteration of this algorithm, a subproblem is

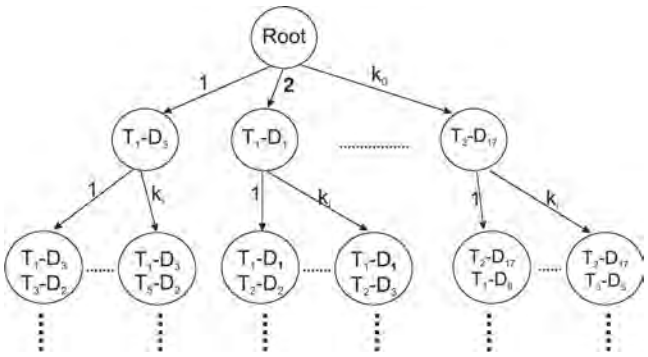


Fig. 2. An example of the truncated greedy algorithm.

selected for exploration from the pool of *live* subproblems using the scores of the current match. We use here a strategy which is similar to the breadth first search strategy found in the literature. The basic principle is to process all the nodes at one level of the search tree before any node at a deeper level.

In what follows, we consider the branching rule for the selected subproblem. Each subproblem is developed by adding one pair to its parent problem. Topology is the most important factor considered in this step. Those (template, data graph) node pairs which are qualified to be added should be connected respectively to at least one template and one data graph node in the parent problem. The way in which a newly added template node connects with the existing template node must be exactly the same as the way that connects the data graph nodes in the corresponding pairs. When a new data graph node is added, the score at that level is calculated using the average of the scores of all the node pairs and edge pairs connecting them, at that level. So as the level increases the new nodes are added and the score is revised in correspondence with the new pair score.

Note that each subproblem at level i ($i = 0, 1, 2, \ldots$) has exactly i pairs (one data graph node and one template node). In summary, the more pairs exist in a subproblem, the more neighboring nodes we have to consider. That causes the search tree to explode exponentially as it goes deeper. In order to avoid such a problem, we do not consider all the feasible pairs. We use a parametric mechanism to control the state space growth. For a subproblem at level i, we only choose at most k_{i+1} best child subproblems. In our work, we first set k_0 to be a fairly large number. The rationale is to make the starting points cover the data graph as much as possible. And then we set $k_i = k$ ($\forall i \geq 1$) with k being relatively small in order to reduce the exponential growth. However the search tree still remains very large. For instance, at level i, the number of sub-problems is given by $k_0 k^{i-1}$, which is extremely large even if the values of k and i are relatively small. Therefore, we introduce two additional parameters β_i and δ. The parameter β_i is to control the breadth of the search tree, i.e., the total number of subproblems at each level is at most β_i. The parameter δ is to control the depth of the search tree, i.e., the search progress stops at level δ with only part of the template explored. If two or more sub-problems at the same level have exactly the same matched pairs, we only retain one of them and fathom the others. It is preferable to let δ be equal to the number of nodes in the template graph for complete exploration.

TABLE I
Ranking of 1-Hop Matched Values

T_1		T_2		T_3		T_4	
D_3	0.90	D_1	0.80	D_6	0.77	D_{10}	0.60
D_1	0.80	D_5	0.75	D_1	0.75	D_7	0.45
D_9	0.75	D_3	0.71	D_2	0.63	D_2	0.40
D_6	0.73	D_7	0.68	D_4	0.52	D_6	0.35
D_5	0.70	D_9	0.55	D_8	0.44	D_4	0.28
D_4	0.65	D_6	0.53	D_9	0.27	D_8	0.18
D_2	0.40	D_4	0.23	D_3	0.23	D_1	0.15
D_7	0.17	D_2	0.21	D_7	0.17	D_3	0.14
D_{10}	0.15	D_{10}	0.13	D_{11}	0.11	D_5	0.12
D_{11}	0.11	D_8	0.09	D_5	0.09	D_9	0.10
D_8	0.09	D_{11}	0.05	D_{10}	0.05	D_{11}	0.07

After running this algorithm, each branch yields a series of matched pairs. Then the data graph nodes in the matched pairs form a subgraph, which is a final match for the template. There are a bunch of such resulting subgraphs with various matched values and topologies, referred to as *leaf nodes*. At each level in algorithm the best β_i leaf nodes are selected. If any of the leaf nodes cannot be extended at level i, then there are no adjacent nodes in the data graph corresponding to the template graph. So a penalty is added to the node and tree is expanded with some non-adjacent node having lower 1-hop neighborhood score. The worst-time complexity of full enumeration of the algorithm is $O(mn^m)$, but in terms of user parameters the complexity is $O(k_0\beta^\delta)$.

We will use the example shown in Fig. 1 to illustrate the TruST algorithm. Table I shows the 1-Hop neighborhood scores calculated by the linear assignment problem. For each node in the template, all nodes in the data graph are ranked in accordance with the corresponding matched values. A higher rank denotes a more similar pair concerning the two nodes themselves and their neighbors. There are a total of 44 matching pairs with threshold, $t = 0$.

Before running the TruST algorithm, we need to determine the values of the parameters. We first set $k_0 = 3$, and then set $k_i = 3$ ($\forall i \geq 1$). For illustrative purposes the value of k_0 is set at the same value as k_i. Since the number of template nodes is 4, we just set $\delta = 4$ to explore all the template nodes. Finally, we set $\beta_i = 7$ ($\forall i \geq 1$).

The results are shown in Fig. 3. At the top level the root node is branched k_0 times. Here only the top three subproblems are selected for further exploration. At each subsequent level each node pair is branched

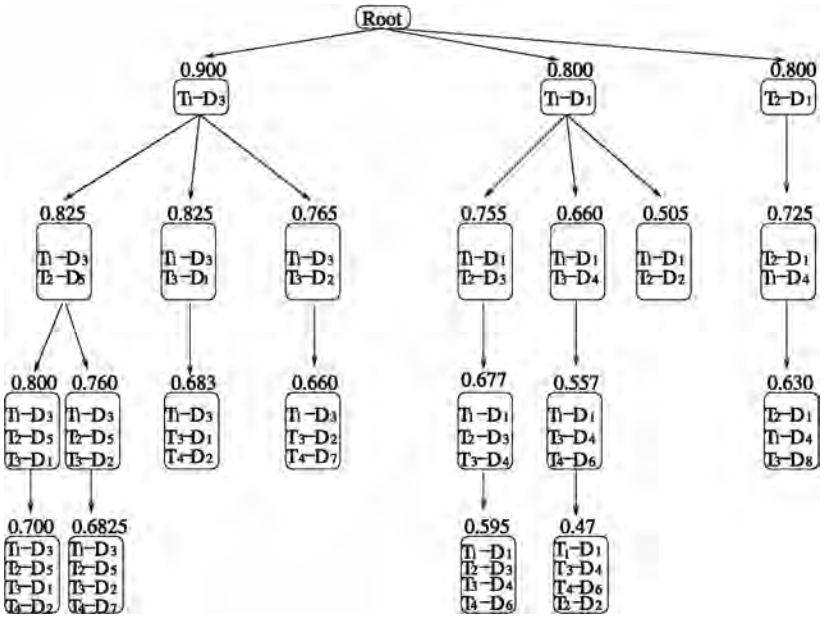


Fig. 3. A sample result of the truncated greedy algorithm.

k_i times. After level 1 we get lot of subproblems, but we only have a choice to continue with β subproblems. After δ level there are no more template nodes left to explore, so at the bottom level in Fig. 3 we find four resulting matches.

4. K-MEANS CLUSTERING USING HYPERCUBE DISTANCE MEASURE

Despite the truncation parameters in the search procedure, it is possible that the user/analyst can be overwhelmed with the number of matched sub-graphs. It is therefore desirable to group subgraphs with similar feature into a fewer number of aggregates. To determine the similarity between a pair of matched subgraphs, we employ the Hypercube distance metric between graphs. Therefore one can use a K-means clustering [33], [23] algorithm to group similar subgraphs. First we try to represent graphs as points in a unit Hypercube. Let M be a score (or 1-hop neighborhood score) matrix for a result graph $R(G)$, with entries in the interval $[0,1]$, where each odd entry M_{ij} indicates a strength of connection between nodes i and j, where $i \in V_{DG}$; $j \in V_{TG}$ and each even entry indicates a strength of connection between edges i and j, where $i \in E_{DG}$; $j \in E_{TG}$. In general, matrices of this type can represent graphs as a whole with neighborhood scores between associated template and graph nodes.

The Hypercube graph representation is obtained by mapping the elements of M into a higher dimensional space, namely, the unit Hypercube of dimension equal to $N = (\min(\delta, m)) \times (\min(\delta, m) - 1)$ for the results of the TruST algorithm, as illustrated in Fig. 4. Of necessity, this mapping induces an ordered indexing correspondence between the matrix elements M and the elements of a vector A in this higher dimensional space.

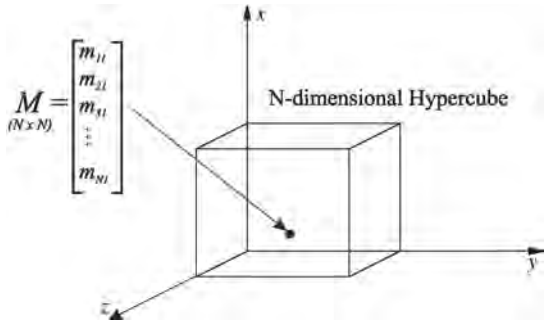


Fig. 4. Embedding a connection matrix into a higher-dimensional unit Hypercube.

The indexing convention we employ to relate the elements A_k of the vector A with the elements $R_{i,1}$ of the matrix R is as follows:

$$A_1 \leftrightarrow R_{1,1}$$
$$\vdots$$
$$A_N \leftrightarrow R_{N,1}$$

or, relating the index k to the row, column pair $(i,1)$ of R,

$$k \leftrightarrow i, \quad i = 1, \ldots, N. \tag{5}$$

Thus the graph represented by the result now becomes a point A in the N dimensional Hypercube, analogous to a set with corresponding coordinates in each of the latter dimensions [28]. A is a set defined on a universe X, so for universe with only one element, the function is defined on a *unit interval* $[0,1]$. For a two-element universe thefunction is defined on the *unit square*; and for a three-element universe, the function is defined on

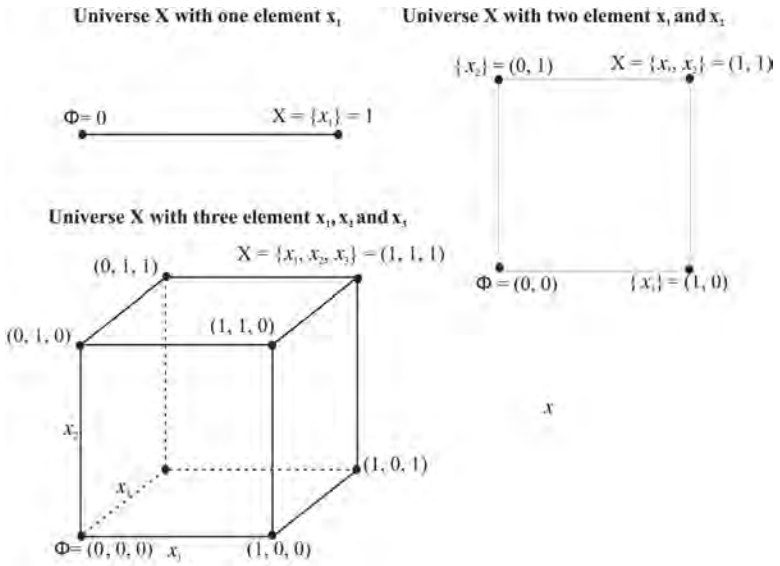


Fig. 5. Graphs as points in the unit hypercube.

the *unit cube*. For a universe of n elements, we define the function on the *unit Hypercube*, $I^n = [0,1]^n$. These mappings are shown in Fig. 5. This mapping can obviously be extended to the different cases of graphs where the total number of elements is different. This Hypercube description allows us to invoke set theoretic concepts for graph representation and characterization.

4.1. Mutual Subsethood of Graphs A and B

Mutual subsethood $E(A,B)$ provides a normalized similarity measure between two graphs. $E(A,B)$ measures the degree to which a set A is similar to another set B. This can be viewed as the degree to which A is a subset of B, AND B is a subset of A. This obviously symmetric relationship is defined by:

$$E(A,B) = \frac{\mu(A \cap B)}{\mu(A \cup B)} \qquad (0 \le E(A,B) \le 1) \qquad (6)$$

where $\mu(A)$ is the Hamming norm of the function values $m_A(y_i)$ of set A:

$$\mu(A) = \sum_{i=1}^{n} m_A(y_i) \qquad (7)$$

and the union (or intersection) operator invokes the component wise maximum (or minimum) operation. Geometrically, mutual subsethood can be visualized as in Fig. 6, as the ratio of the Hamming lengths of two vectors, the numerator vector having as its coordinates the element-wise minima of A and B, while the denominator vector has as it coordinates the element-wise maxima of A and B.

Note that $E(A,B) = 1 \Leftrightarrow A = B$, and $E(A,B) = 0$ if A or $B = \phi$ where ϕ denotes the null set at the origin of I^n. The mutual subsethood measure can also straightforwardly incorporate dimensional importance weighting.

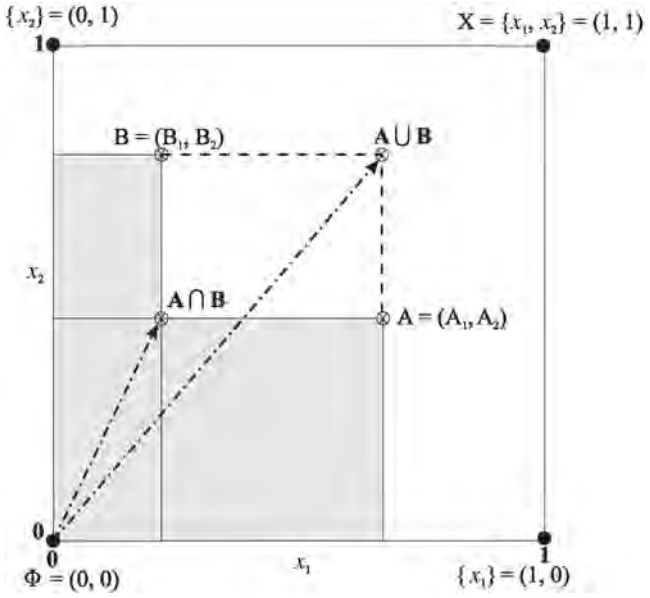


Fig. 6. Geometric interpretation of mutual subsethood as the ratio of Hamming lengths of the two vectors shown.

4.2. Hypercube Distance between Graphs A and B

The most important benefit of this Hypercube representation is that it immediately suggests the use of a simple function of mutual subsethood as a distance metric between pairs of labeled graphs. Since $E(A,B) = 1 \Leftrightarrow A \equiv B$ and $E(A,B) = 0$ if and only if A and B share no links in common, the normalized distance metric $\Delta(A,B)$ between two arbitrary labeled graphs can be defined as the complement of $E(A,B)$:

$$\Delta(A,B) = 1 - E(A,B), \qquad (0 \le \Delta(A,B) \le 1). \quad (8)$$

The distance achieves a maximum of unity if A and B have no common links. In this case, we define the indeterminate ratio in (6) to have zero value, so that the distance between two such graphs as defined by (8) is

unity. This avoids the introduction of a discontinuity in the value of $\Delta(A,B)$ in the limit as the number of edges in a pair of statistically independent graphs A and B approaches zero. Note that the computation of $\Delta(A,B)$ is very fast, even for large graphs, since it involves only pair wise comparator calculations within each dimension, as opposed to the more numerically sensitive and computationally intensive calculations involved in the eigen analysis of a graph.

Clearly, $\Delta(A,B) \geq 0$ and $\Delta(A,B) = 0 \Leftrightarrow A = B$. We now prove that $\Delta(A,B)$ is a proper metric, i.e., that it also satisfies the triangle inequality for distinct graphs A, B and C:

$$\Delta(A,C) \leq \Delta(A,B) + \Delta(B,C). \tag{9}$$

From (6), we have

$$\begin{aligned}\Delta(A,C) &= 1 - \frac{\sum_{k=1}^{N(N-1)/2} \min(A_i,C_i)}{\sum_{i=1}^{N(N-1)/2} \max(A_i,C_i)} \\ &\leq 1 - \frac{\sum_{k=1}^{N(N-1)/2} \min(A_i,C_i)}{\sum_{i=1}^{N(N-1)/2} \max(A_i,B_i,C_i)} \\ &= \frac{\sum_{i=1}^{N(N-1)/2} \max(A_i,B_i,C_i) - \sum_{k=1}^{N(N-1)/2} \min(A_i,C_i)}{\sum_{i=1}^{N(N-1)/2} \max(A_i,B_i,C_i)}.\end{aligned} \tag{10}$$

Similarly, we have Equation (11)

$$\begin{aligned}&\Delta(A,B) + \Delta(B,C) \\ &\quad= \frac{\sum_{i=1}^{N(N-1)/2} \max(A_i,B_i) - \sum_{i=1}^{N(N-1)/2} \min(A_i,B_i)}{\sum_{i=1}^{N(N-1)/2} \max(A_i,B_i)} \\ &\qquad+ \frac{\sum_{i=1}^{N(N-1)/2} \max(B_i,C_i) - \sum_{i=1}^{N(N-1)/2} \min(B_i,C_i)}{\sum_{i=1}^{N(N-1)/2} \max(B_i,C_i)} \\ &\quad\geq \frac{\sum_{i=1}^{N(N-1)/2} \max(A_i,B_i) + \sum_{i=1}^{N(N-1)/2} \max(B_i,C_i) - \sum_{i=1}^{N(N-1)/2} \min(A_i,B_i) - \sum_{i=1}^{N(N-1)/2} \min(B_i,C_i)}{\sum_{i=1}^{N(N-1)/2} \max(A_i,B_i,C_i)}.\end{aligned} \tag{11}$$

Since the latter terms in both (10) and (11) have a common denominator, it suffices to show that the numerators of these terms satisfy the inequality

$$\begin{aligned}&\sum_{i=1}^{N(N-1)/2} \max(A_i,B_i,C_i) - \sum_{k=1}^{N(N-1)/2} \min(A_i,C_i) \\ &\quad\leq \sum_{i=1}^{N(N-1)/2} \max(A_i,B_i) + \sum_{i=1}^{N(N-1)/2} \max(B_i,C_i) \\ &\qquad- \sum_{i=1}^{N(N-1)/2} \min(A_i,B_i) - \sum_{i=1}^{N(N-1)/2} \min(B_i,C_i).\end{aligned} \tag{12}$$

This inequality can be demonstrated straightforwardly to hold for each value of the index i by considering each of the six possible rankings of the magnitudes of A_i, B_i and C_i, and thus the triangle inequality holds for the distance metric $\Delta(A,B)$.

The graph matching example solved using the TruST algorithm results in matches shown in Fig. 3. Using this distance metric, we treat graphs as vector samples, upon which vector processing operations can be performed. In our example we have four resulting subgraphs and so we can create $(\beta - 1) = 3$ clusters. The Hypercube distance measure is used to calculate K-means cluster. Based on the distances the leaf nodes are distributed to different clusters. The whole process continues until the clusters rearrangement stops. Cluster 1 and 2 each have only one leaf node while cluster 3 has two leaf nodes. Here the clusters formed are:

$$\begin{aligned}&\text{Cluster 1}: \quad DP_1 \\ &\text{Cluster 2}: \quad DP_2 \\ &\text{Cluster 3}: \quad DP_3, DP_4.\end{aligned}$$

5. K-MEANS CLUSTERING USING FUZZY HAMMING DISTANCE MEASURE

In this type of representation, the graphs are represented as vectors in Hypercube space. This method of graph representation is similar to the one used in Section 4 but Fuzzy Hamming distance measure is used instead of the mutual subsethood. Each graph in the result set is represented as a vector shown by Equation 6.

Given two real-valued vectors, FHD is the (fuzzy) number of components along which the two vectors are different. Degree of difference between vectors is defined as "Given the real values x and y, the degree of the difference between x and y, modulated by $\alpha > 0$, denoted by $d_\alpha(x,y)$, is defined as:"

$$d_\alpha(x,y) = 1 - e^{-\alpha(x-y)^2}. \tag{13}$$

For the same value of $(x-y)$, α will vary the value of $d_\alpha(x,y)$. The membership function d_α has following properties:

1. $0 \leq d_\alpha(x,y) < 1$ with equality $\Leftrightarrow x = y$;
2. $d_\alpha(x,y) = d_\alpha(y,x)$;

3. for $x = a \pm c$, $d_\alpha(x,a) = e^{-c^2}$;
4. $d_\alpha(x,y) = d_\alpha(0,|x-y|)$.

The first three points above prove that FHD is a metric. The *difference fuzzy set* corresponding to the $d_\alpha(x,y)$ is given as $D_\alpha(x,y)$ with membership function $\mu_{D_\alpha(x,y)}$ such that:

$$\mu_{D_\alpha(x,y)}(i) = d_\alpha(x_i, y_i) \tag{14}$$

where x_i and y_i are the ith component in n dimensional vector x and y respectively. The difference fuzzy set gives the measure of degree of difference between vectors along their ith component. Ionescu and Ralescu [19] define FHD as:

> "Given two n dimensional real-valued vectors, x and y, for which the difference fuzzy set $D_\alpha(x,y)$, with membership function $\mu_{D_\alpha(x,y)}$, the **Fuzzy Hamming Distance** between x and y, denoted by $\mathrm{FHD}_\alpha(x,y)$ is the fuzzy cardinality of the difference fuzzy set, $D_\alpha(x,y)$."

Cardinality Card$A$ of a fuzzy set $A = \sum_{i=1}^{n} x_i/\mu_i$:

$$\mathrm{Card}\, A \equiv \sum_{i=0}^{n} i/\mu_{\mathrm{Card}\, A}(i) \tag{15}$$

where $\mu_{\mathrm{Card}\, A}(i) = \min(\mu_i, (1-\mu_{i+1}))$. μ_i denotes the ith largest value of μ. The set is appended with 1 at start and 0 at end for convenience of calculation. So, $\mu_0 = 1$ and $\mu_{n+1} = 0$.

These values are the fuzzy equivalent of the graph vector. To get a distance measure the fuzzy values are to be defuzzified to get a crisp distance value. The literature on fuzzy inference system has many different types of defuzzification techniques. The most used among these is the "*Center of Gravity* (COG)." Ionescu and Ralescu [19] have proposed a crisp cardinality (nCard) to get a integer defuzzified value

$$n\mathrm{Card}(A) \equiv |\overline{x; \mu_A(x) > 0.5}|. \tag{16}$$

The COG is defined as per its name, as the center of gravity of the area underlying the resulting graph of Fuzzy Hamming set and its membership function values. It gives the number of different components that exists between vectors x and y. nCard gives the integer number of components that are different in the fuzzy set. The COG is given by the equation:

$$\mathrm{COG} = \frac{\sum_{i=1}^{n} x_i \mu_i}{\sum_{i=1}^{n} \mu_i}. \tag{17}$$

The main issue in this distance measure is calculation of α. α can vary the measure of the FHD. α acts as a control for sensitivity of variation. It is given as:

$$\alpha = \frac{\ln\left(\frac{1}{\epsilon}\right)}{\mathrm{MAX}^2} \frac{1}{\beta^2} \tag{18}$$

where MAX is the maximum value in the column domain and β is the percentage of MAX which is considered a difference in the column values of vectors x and y. In our case the graph vectors have values in the range $[0,1]$, so the maximum difference is going to be 1. Hence MAX = 1. The nCard decreases significantly as we increase β. So to have a more sensitive analysis of difference the value of β is kept at 0.1.

5.1. Implementation

We again use the same k-means clustering but with the distance measure as Fuzzy Hamming distance. The graph matching example solved using the TruST algorithm results in matches shown in Fig. 3. In our example we have four resulting subgraphs and so we can create $(\beta - 1) = 3$ clusters. For K-means clustering the 3 cluster centroids are initialized with the first three results.

$$\begin{aligned} K_1 &= (0.9, 0.80, 0.75, 0.75, 0.75, 0.63, 0.40) \\ K_2 &= (0.9, 0.80, 0.75, 0.71, 0.63, 0.59, 0.45) \\ K_3 &= (0.8, 0.75, 0.71, 0.68, 0.52, 0.55, 0.35). \end{aligned} \tag{19}$$

Then Fuzzy Hamming distance is calculated from cluster centroids to each leaf node (LN). The FHD is calculated using Equation 16. Here a sample FHD calculation between cluster 1 centroid and Leaf Node 2 is shown. From the earlier discussion we set MAX as 1, β as 0.1 and ϵ as 0.5. Therefore, $\alpha = 69.31$.

$$\begin{aligned} d_\alpha(x,y) &= 1 - e^{-69.31 \times (x-y)^2} \\ d_{69.31}(x_1, y_1) &= 1 - e^{-69.31 \times (0.90-0.90)^2} = 0 \\ d_{69.31}(x_2, y_2) &= 1 - e^{-69.31 \times (0.80-0.80)^2} = 0 \\ d_{69.31}(x_3, y_3) &= 1 - e^{-69.31 \times (0.75-0.75)^2} = 0 \\ d_{69.31}(x_4, y_4) &= 1 - e^{-69.31 \times (0.75-0.71)^2} = 0.1045 \\ d_{69.31}(x_5, y_5) &= 1 - e^{-69.31 \times (0.75-0.63)^2} = 0.6310 \\ d_{69.31}(x_6, y_6) &= 1 - e^{-69.31 \times (0.63-0.59)^2} = 0.1050 \\ d_{69.31}(x_7, y_7) &= 1 - e^{-69.31 \times (0.40-0.45)^2} = 0.1590 \end{aligned} \tag{20}$$

$$n\mathrm{Card}(A) \equiv |\overline{x; \mu_A(x) > 0.5}| = 1. \tag{21}$$

Based on the distances the leaf nodes are distributed to different clusters. The whole process continues until the clusters rearrangement stops. Here the clusters formed are:

$$\begin{aligned} &\text{Cluster 1}: \quad DP_1, DP_2, DP_3 \\ &\text{Cluster 2}: \quad — \\ &\text{Cluster 3}: \quad DP_4. \end{aligned}$$

Cluster 1 has three data points, while cluster 3 has one data point. Cluster 2 is empty in this case. This clustering method groups graphs differently than the Hypercube based k-means clustering, where cluster 1 and 2 each have only one leaf node while cluster 3

TABLE II
Silhouette Indices

	FH	HYP	SA	Silhouette Index
FH	S(FH,FH)	S(FH,HYP)	S(FH,SA)	(S(FH,FH)+S(FH,HYP)+S(FH,SA))/(3)
HYP	S(HYP,FH)	S(HYP,HYP)	S(HYP,SA)	(S(HYP,FH)+S(HYP,HYP)+S(HYP,SA))/(3)
SA	S(SA,FH)	S(SA,HYP)	S(SA,SA)	(S(SA,FH)+S(SA,HYP)+S(SA,SA))/(3)

has two leaf nodes. Analysis of these measures is done in next section, to see which distance measure is more efficient.

6. NUMERICAL TESTING

Cluster validation is a very important issue in clustering analysis because the result of clustering needs to be validated in most cases. It is defined as measuring goodness of a clustering relative to others created by the same algorithms using different parameter values. In most clustering algorithms, as k-means, the number of clusters is set as user parameter. There are many approaches toward cluster validation. One of such techniques which we use is the "*Silhouette validation technique*."

The Silhouette validation technique [29] calculates the silhouette width for each sample, average silhouette width for each cluster and overall average silhouette width for a total data set. Using this approach each cluster could be represented by a so called silhouette, which is based on the comparison of its tightness and separation. The average silhouette width could be applied for evaluation of clustering validity and can also be used to decide how good is the number of selected clusters. To construct the silhouettes $S(i)$ the following formula is used:

$$S(i) = \frac{(b(i) - a(i))}{\max(a(i), b(i))} \tag{22}$$

where $a(i)$ is the average dissimilarity of i object to all other objects in the same cluster; $b(i)$ is the minimum of average dissimilarity of i object to all objects in other cluster (in the closest cluster).

It is followed from the formula that $-1 \leq S(i) \leq 1$. If silhouette value is close to 1, it means that sample is "well clustered" and it was assigned to a very appropriate cluster. If silhouette value is near zero, it means that the sample could be assigned to another closest cluster as well, and the sample lies equally far away from both clusters. If silhouette value is close to -1, it means that sample is "misclassified" and is merely somewhere in between the clusters. The overall average silhouette width for the entire plot is simply the average of the $S(i)$ for all objects in the whole dataset. The largest overall average silhouette indicates the best clustering. Therefore, the number of cluster with maximum overall average silhouette width is taken as the optimal number of the clusters.

Here we have two distance measures Hypercube distance (HYP) and Fuzzy Hamming distance (FH). We also used a simple average distance measure (SA) which is independent of template graphs. The HYP and FH distance are bound by the axes which are defined by the template nodes and edges to which the nodes and edges in data graph are matched. In SA distance the K-medoid algorithm is used. The distance between two matches is calculated using the average of data graph node-to-node scores (DGN) and data graph edge-to-edge scores (DGE). In HYP and FH distance the data graph node (edge) to template graph node (edge) scores are used. Let's say there are two matches for a template graph $(T_1 - E_1^T - T_2 - E_2^T - T_3)$ given as $(D_1 - E_1^D - D_2 - E_2^D - D_3)$ and $(D_4 - E_3^D - D_5 - E_4^D - D_6)$. Then the SA distance is given as:

$$SA = \frac{\begin{array}{l}(DGN(D_1, D_4) + DGE(E_1^D, D_3^D) \\ \quad + DGN(D_2, D_5) + DGE(E_2^D, E_4^D) + DGN(D_3, D_6))\end{array}}{5}.$$

Now we have used three distances for clustering and we have three Silhouette indices for each of them. Each clustering algorithm is analyzed using all the three distance measures. The Silhouette index for each clustering is calculated as an average (Table II) of the three Silhouette indices from the three distance measures. After generating 100 random sample runs for each of the clustering algorithms, we found that, only 2 times SA performed better or equal to FH and HYP. So we will just compare FH and HYP based clustering. But the Silhouette index from SA is used during analysis.

The two methods are independent of each other so the best way to compare them is to conduct a statistical analysis. A "2 sample t-test" is planned for comparing the two distances. A power and sample size capability analysis is conducted to evaluate sample size before we design and run an experiment. It tells us the right amount of runs required to detect a notable difference during the statistical analysis. After conducting some sample runs we have the data with standard deviation of 0.27. The type 1 error (α) is set at 0.05 and type 2 error (β) at 0.2. These values are used to determine the sample size. To detect a difference of 0.1 we require 116 runs and for 0.15 we require 55 runs. To detect the difference accurately, the experiment was decided to be conducted on 100 runs.

T-test requires data to be normally distributed and the samples to have equal variance. To check for the assumptions, normality test and F-test are conducted.

The normality test for Fuzzy Hamming and Hypercube suggested that the p-values are less than α, therefore we reject H_0, i.e. the data is not normally distributed. The F-test has a null (H_0) hypothesis that the variances of both the methods are equal and alternative (H_1) hypothesis that they are not equal. If the p value is greater than α then we fail to reject H_0 and else we reject H_0. Here the p value is greater than α, so the variances of both the distance measures are equal.

The assumption of equal variances is satisfied, while the test for normality is rejected, so t-test cannot be conducted. The data is not normal, so a non-parametric test is conducted. There are a number of non-parametric tests that can be used to test the difference between two non-normal sets of sample data. In those cases, we are testing the difference in medians, and not means. Mann-Whitney test is the nonparametric test equivalent to the 2 sample t-test. In Mann-Whitney test the hypotheses are:

$$H_0: \quad \eta_1 = \eta_2$$

$$H_1: \quad \eta_1 \neq \eta_2, \quad \text{where } \eta \text{ is the population median.}$$

The sample medians of the ordered data are 0.2597 and 0.3334. The 95.0% confidence interval for the difference in population medians ($\eta_1 - \eta_2$) is [−0.0627 to 0.0737]. The test statistic $W = 10131$ has a p-value of 0.8441 when adjusted for ties. Since the test is significant at 0.8441, we conclude that there is sufficient evidence to reject H_0. Therefore, the data supports the hypothesis that there is a difference between the population medians. Now the two-sided hypothesis is changed to one-sided hypothesis comparing the medians. Looking at the medians for Fuzzy Hamming (FH) and Hypercube (HYP) we change the hypotheses to:

$$H_0: \quad \eta_1 = \eta_2$$

$$H_1: \quad \eta_1 < \eta_2.$$

The test is significant at 0.4220. This implies that the null hypothesis is rejected only if median of HYP is larger than the median of FH. Based on this result we can say that Hypercube based K-means clustering performs better than Fuzzy Hamming based K-means clustering. For those runs where FH performed better than HYP, the difference between the Silhouette indices was less than 13% for 75% of the runs.

7. AGGREGATING CLUSTERS

Using the Truncated Search Tree algorithm, we got multiple matches for the template. The matches were clustered using K-means clustering algorithm. For K-means clustering we experimented with two distance measures "Hypercube distance" and "Fuzzy Hamming distance." Section 6 shows that "Hypercube distance" performed better than "Fuzzy Hamming distance." Now to find the important nodes and edges in the formed cluster we aggregate the matches. The aggregation helps reduce the analyst's area of concentration to a small region of data graph.

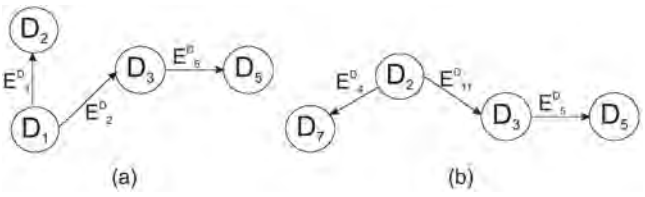


Fig. 7. Aggregation results. (a) Cluster 1. (b) Cluster 2.

TABLE III
Cluster 3

DP_3	DP_4
D_1	D_1
E_2^D	E_1^D
D_3	D_2
E_3^D	E_3^D
D_4	D_4
E_6^D	E_6^D
D_6	D_6

To aggregate the clusters we combine the matches using union and intersection operation. Union operation is a way of diversification zooming on the data graph while intersection operation is a way of intensification zooming. Union of clusters gives us a more diverse information to concentrate on, while intersection gives us the most important information. In the union operation all the unique nodes and edges in the cluster are marked significant. In the intersection operation the common nodes and edges between the clustered matches are marked significant. In this section an example is presented to show the union and intersection of clusters. The example previously used in this paper will be used for further aggregation.

As Section 6 shows that "Hypercube distance" performed better than "Fuzzy Hamming distance," we will use the clusters formed using "Hypercube distance" for further aggregation. As clusters 1 and 2 have only one result node so there won't be any aggregation and the results will be displayed as a whole match. The aggregation results for cluster 1 and 2 are shown in Fig. 7(a). and 7(b) respectively. But in case of cluster 3 there are two result nodes DP_3 and DP_4 (Shown in Table III).

Using the intersection method of aggregation, we obtain the nodes and edges shown in Table V. We can see that nodes D_1, D_4 and D_6 and edges E_3^D and E_6^D are common to both result nodes. The nodes and edges obtained using union method of aggregation are shown in Table IV. In addition to nodes and edges in Table V, we have nodes D_2 and D_3 and edges E_1^D and E_2^D that are uncommon in the two result nodes.

The aggregation results for union and intersection method in Cluster 3 are shown in Fig. 8(a) and 8(b) respectively.

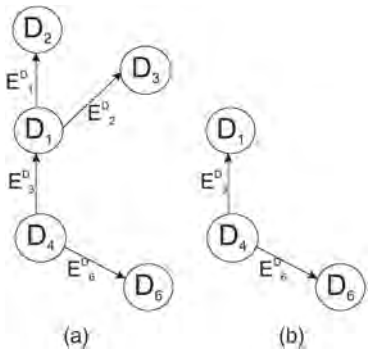


Fig. 8. Cluster 3 aggregation result. (a) Union. (b) Intersection.

TABLE IV
Cluster 3 Union

Nodes	Edges
D_1	E_1^D
D_2	E_2^D
D_3	E_3^D
D_4	E_6^D
D_6	

TABLE V
Cluster 3 Intersection

Nodes	Edges
D_1	E_3^D
D_4	E_6^D
D_6	

8. NEIGHBORHOOD INFORMATION

"Neighborhood" is a word with many different levels of meaning in mathematics. One of the most general concepts of the neighborhood of a point x in (R^n) ϵ-neighborhood or infinitesimal open set, is the set of points inside an n-ball with center x and radius $\epsilon > 0$. The neighborhood of a vertex v in a graph is the set of all the vertices adjacent to v. More generally, the ith neighborhood of v is the set of all vertices that lie at the distance i from v.

In graph theory, the neighborhood of a vertex v in a graph G is the induced subgraph of G consisting of all vertices adjacent to v and all edges connecting two such vertices. Two vertices u and v are considered adjacent if an edge exists between them. This is denoted by $u \downarrow v$. The neighborhood [34] is often denoted $N_G(v)$ or (when the graph is unambiguous) $N(v)$. The same neighborhood notation may also be used to refer to sets of adjacent vertices rather than the corresponding induced subgraphs. The neighborhood described above does not include v itself, and is more specifically the open neighborhood of v; it is also possible to define a neighborhood in which v itself is included, called the closed neighborhood and denoted by $N_G[v]$. When stated without any qualification, a neighborhood is assumed to be open.

If all vertices in G have neighborhoods that are isomorphic to the same graph H, G is said to be locally H, and if all vertices in G have neighborhoods that belong to some graph family F, G is said to be locally F. Neighborhoods are also used in the clustering coefficient of a graph, which is a measure of the average density of its neighborhoods. In [16], authors have presented a neighborhood broadcast and gossiping problem. For neighborhood broadcast the authors have considered nodes which are one edge away and for gossiping problem they have considered nodes, which are accessible from the main node. Schenker *et al.* [32] have compared a vector based graph representation, combined with a k-Nearest Neighbor (k-NN) algorithm to the graph matching approach, to represent web documents. There is a limited amount of literature available in the field of neighborhood structure, but we have found the concept of neighborhood of nodes used in various fields to calculate distances, get density of graphs, etc.

We have applied the truncated branch and bound method to match the templates to data graphs. Based on the settings for the algorithm parameters, we get various matches for the given template. The matches are clustered using K-means clustering, to find the most significant information among them. But this information is limited to the template we are trying to match. Consider a terrorist template which we are trying to match against a given social network represented as data graph. We can find the most plausible terrorist networks in the given data graph. Having this information we have no idea as to what the terrorists are planning. So the neighborhood of a match is as important as the match. To get more information about the neighborhood of the matches we have developed an algorithm to find the neighborhood score for each node in data graph.

8.1. Problem Definition

8.1.1. Neighborhood Structure of Matches

The data graph represented as G_D can be divided into two sets. The first set is the match of the template graph in the data graph, called "*Core graph*," C_D. The other set contains the remaining nodes in graph G_D called periphery nodes $(G_D - C_D)$. Now to find the most relevant neighborhood structure, we need to find the neighborhood scores for the periphery nodes with respect to the match. The neighborhood score must represent the connectivity of the periphery nodes to the core nodes. In the case of terrorist networks, all the members have a single aim and they will concentrate their energies toward that target. To reach the target, members will use covert routes to hide their intent. To find these targets we need the most accurate possible neighborhood information. This score can be calculated

using this formula:

$$P_i = \frac{\sum_j d_{ij}}{N_T}.$$

Here P_i is the periphery node score for node i, N_T is the number of nodes in the match, while d_{ij} is the distance between periphery node i and core node j. d_{ij} is the distance measure between core and periphery nodes. This distance is the weighted shortest path distance divided by the length of the path. As the edge weights lie in the interval $[0,1]$, the distance measure, d_{ij}, also lies in the interval $[0,1]$. Therefore, $(0 \leq P_i \leq 1)$

The neighborhood score takes into consideration the distance and connectivity of the periphery node to the match. The distance d_{ij} is the measure of importance of node i with respect to node j in C_D, i.e. a relationship between the possible target and the member of the terrorist cell. Hence, P_i helps to find the probability that node i is the most common target sought by all the members in the group (match).

In some cases the neighborhood structure can contain a person or contact of interest and the contact may not be accessed by all the members. The contact can be a very vital link between two groups. To find this type of information in neighborhood we need to find the nodes which have the maximum weight connectivity to the core nodes. The neighborhood score is calculated using this formula:

$$P_i = \max_j(d_{ij}).$$

The problem of finding relevant neighborhood information can be also viewed in the sense of maximum weight connectivity and here it is considered out of scope, and hence won't be considered for analysis.

8.1.2. Ranking Matches based on Neighborhood Scores

After finding the neighborhood structure, we have a lot of information which is scattered around the matches. The matches by themselves have a ranking mechanism, but that is independent of the neighborhood structure they represent. The neighborhood structure may be important for a match with a low ranking score. Therefore, we would like to find out which match has the highest neighborhood score. To rank the matches based on neighborhood structure, we take the average of the neighborhood scores for all the nodes in periphery graph $(G_D - C_D)$. Using this new score, we will rank the matches and this will signify the importance of matches based on the profile of the target they are attacking.

8.1.3. Neighborhood Structure of Clusters

After calculating the neighborhood structure for the matches, we want to find the most important structure around the clusters since this is the aggregated result presented to the analyst. Now with these aggregated results, the neighborhood information will be different.

TABLE VI
Edge Weights

Edge	From Node	To Node	Weights
E_1^D	D_1	D_2	0.488388
E_2^D	D_1	D_3	0.886889
E_3^D	D_4	D_1	0.268496
E_4^D	D_2	D_7	0.774886
E_5^D	D_3	D_5	0.046768
E_6^D	D_4	D_6	0.437913
E_7^D	D_8	D_4	0.427433
E_8^D	D_4	D_9	0.361861
E_9^D	D_{11}	D_5	0.211624
E_{10}^D	D_6	D_{10}	0.772303
E_{11}^D	D_2	D_3	0.277781
E_{12}^D	D_9	D_{10}	0.090555
E_{13}^D	D_5	D_{10}	0.486635

To calculate the neighborhood of a cluster we take the data graph represented as G_D and divide it into two sets. The cluster formed using the matches of the template graph in the data graph, represents C_D. Now to find the most relevant neighborhood structure, we need to find the neighborhood scores for the periphery nodes with respect to the cluster. This neighborhood information will be different from the information we get in Section 8.1.1. In Section 8.2 an illustrative example is presented to show the calculation for neighborhood structures of matches, clusters and ranking of matches.

8.2. Implementation

For the given data and template graph, we first find the weights for the edges in data graph. The weights for all the edges in data graph are given in Table VI. Using the edge weights, the shortest path distance between all the nodes in the data graph is calculated.

To calculate the neighborhood structure of the matches we select one of the subgraphs generated by the matching process in Section 3. For illustration purpose let us select match 1. Fig. 9 shows the match with respect to the data graph. In this example the node set $\{D_1, D_2, D_3, D_5\}$; along with edge set $\{E_1^D, E_2^D, E_5^D\}$ represents the core nodes for the given data graph. The rest of the nodes in the data graph represent the periphery nodes and edges for match 1. For match 1, there are 7 neighborhood nodes. For data graph node D_4 we can calculate the neighborhood score as follows:

$$P_4 = \frac{\sum_j d_{4j}}{N_T} = \frac{0.488388 + 0.270395 + 0.178724 + 0.277781}{4}$$

$$= 0.303822.$$

If the problem of finding relevant neighborhood information is viewed in the sense of maximum weight connectivity then match 1 will have nodes D_4, D_6, D_7, D_8, D_9, D_{10} and D_{11} as the periphery nodes. Here node D_2 is connected to node D_7 only. Node D_1 is connected

TABLE VII
Match Neighborhood based on Maximum Edge Connectivity

	D_4	D_6	D_7	D_8	D_9	D_{10}	D_{11}
D_1	0.268496	0.3532045	0	0.3479645	0.3151785	0.492904	0
D_2	0	0	0.774886	0	0	0	0
D_3	0	0	0	0	0	0	0
D_5	0.565617	0.629469	0	0.531071	0.288595	0.486635	0.211624
Neighborhood Score	0.565617	0.629469	0.774886	0.531071	0.3151785	0.492904	0.211624

TABLE VIII
Subgraph Ranking based on Neighborhood Score

Match 4	0.364657
Match 3	0.344861
Match 1	0.332481
Match 2	0.326449

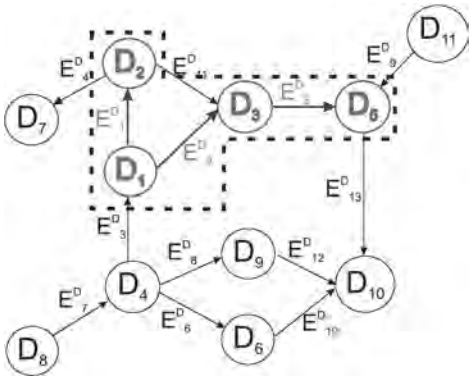


Fig. 9. Match neighborhood structure.

to all the nodes except D_7 and D_{11}. D_3 being a central node has no periphery associated to it. D_5 has node D_7 in the periphery, which is not connected. Based on the edge weights given in Table VI, the neighborhood score for all the periphery nodes with respect to match 1 are shown in Table VII.

After calculating the neighborhood scores for all neighborhood nodes, we would like to rank the subgraph matches based on the importance of neighborhood. To find the neighborhood score for each match we take the average of the neighborhood scores for all the nodes in the periphery of the given match. This is the new match score based on the neighborhood values. Then based on these scores we rank the subgraphs. The matches are ranked as shown in Table VIII.

To calculate the neighborhood structure of the clusters we select one of the clusters generated by the aggregating process in Section 7. For illustration purpose let us select cluster 3, as cluster 1 and 2 have only one leaf node. Cluster 1 and 2 will have a neighborhood structure similar to match 1 and 2 respectively. Fig. 10 shows the cluster union with respect to the data graph. In this example the node set $\{D_1, D_2, D_3, D_4, D_6\}$; along with edge set $\{E_1^D, E_2^D, E_3^D, E_6^D\}$ represents the core nodes for

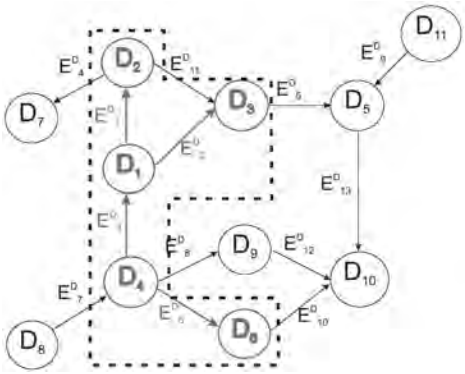


Fig. 10. Cluster 3 (union) neighborhood structure.

the given cluster in data graph. The core node and edge sets are marked in Fig. 10. The rest of the nodes in the data graph represent the periphery nodes and edges for cluster 3 union. For the cluster 3 union, there are 6 neighborhood nodes. Similar to Match neighborhood we calculate the neighborhood scores for Cluster aggregates.

9. CONCLUSION

High level Data Fusion, also recognized as Level 2 (situation assessment) and Level 3 (impact assessment) in the JDL architecture involves understanding relationships between level 1 objects, which can be well modeled using attributed graphs. These graphical models can be applied in a wide range of applications such as, Cyber Security, Asymmetric Warfare, Disease Surveillance, Intelligence and Knowledge Discovery, and Improvised Explosive Device Detection. In these domains a complex situation of interest to analyst can be formulated as a template graph which represents alternative hypothesis. The analyst is interested in determining the occurrence of that situation in a sensor database which is in turn represented as a data graph. This gives rise to an inexact subgraph matching problem, which is NP-Hard and can lead to large number of matches. To overcome the problem associated with the large number of matches, an approach is necessary to fuse the resulting subgraphs. The dimensionality of the resulting inexact matches can be reduced using a clustering algorithm which can quantify the similarity between any two graphs via a selected similarity measure.

We presented a TRUncated Search Tree algorithm [31] to find the appropriate subgraphs of the data graph for the template. Since the state space of a branch-and-bound method explodes as the search tree goes deeper and deeper, this algorithm considers only part of the state space and truncates those with relatively small values. The resulting subgraphs generated using TruST algorithm are then aggregated using the K-means clustering algorithm. The algorithm makes use of Hypercube distance measure. This distance measure finds similarity distance between any two given subgraphs. Then the TruST algorithm is compared with a similar fuzzy distance measure called as Fuzzy Hamming distance metrics. To find the best distance measure Silhouette indices are calculated. The results show that Hypercube based K-means clusters are well separated. Then the Hypercube based K-means clusters are fused using union and intersection based aggregation method. The aggregation of clusters gives analyst a particular location to concentrate his analysis. The neighborhood structure of the matches and aggregated clusters is explored to find significant nodes. These nodes found in the neighborhood structure do not form the initial hypothesis, but can be of real importance to the analyst.

We have applied TruST along with Hypercube based K-means clustering and aggregation, in modeling and simulation of asymmetric warfare situations with encouraging results. The algorithm matches the complexity of the domains of application and user parameter settings. A further insight on how to select the breath and depth control parameters is provided in [31]. Another interesting area of research will be developing a dynamic algorithm for level 2 fusion. In this algorithm the relationship between newly added information and generated clusters will be checked and new aggregated clusters will be developed eventually.

REFERENCES

[1] M. Balazinska, A. Deshpande, M. Franklin, P. Gibbons, J. Gray, S. Nath, M. Hansen, M. Liebhold, A. Szalay and V. Tao
Data management in the worldwide sensor web.
Pervasive Computing, IEEE, **6**, 2 (Apr.–June 2007), 30–40.

[2] D. Ballard and L. Rippy
A knowledge-based decision aid for enhanced situational awareness.
In *IEEE/AIAA 13th Digital Avionics Systems Conference*, 1994, 340–347.

[3] S. B. Banks and C. S. Lizza
Pilot's associate: A cooperative, knowledge-based system application.
IEEE Intelligent Systems and Their Applications, **6**, 3 (1991), 18–29.

[4] E. Bengoetxea, P. Larranaga, I. Bloch and A. Perchant
Estimation of distribution algorithms: A new evolutionary computation approach for graph matching problems.
In *EMMCVPR '01: Proceedings of the Third International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, London, UK: Springer-Verlag, 2001, 454–468.

[5] E. Bengoetxea, P. Larranaga, I. Bloch, A. Perchant and C. Boeres
Inexact graph matching using learning and simulation of bayesian networks.
In *Proceedings of CaNew workshop, ECAI*, 2000.

[6] E. Bengoetxea, P. Larranaga, I. Bloch, A. Perchant and C. Boeres
Inexact graph matching by means of estimation of distribution algorithms.
Pattern Recognition, **35**, 12 (2002), 2867.

[7] A. Bookstein, S. T. Klein and T. Raita
Fuzzy hamming distance: A new dissimilarity measure.
In *CPM '01: Proceedings of the 12th Annual Symposium on Combinatorial Pattern Matching*, London, UK: Springer-Verlag, 2001, 86–97.

[8] H. Bunke and K. Shearer
A graph distance metric based on the maximal common subgraph.
Pattern Recognition Letters, **19**, 3–4 (1998), 255.

[9] G. Chartrand, F. Saba and H. B. Zou
Edge rotation and distance between graphs.
Cas. Pest. Mat. (Mathematica Bohemica), **110** (1985), 87–91.

[10] L. Chin-Teng and C. S. G. Lee
Reinforcement structure/parameter learning for neural-network-based fuzzy logic control systems.
IEEE Transactions on Fuzzy Systems, **2**, 1 (1994), 46–63.

[11] T. Coffman, S. Greenblatt and S. Marcus
Graph-based technologies for intelligence analysis.
Communications of the ACM, **47**, 3 (Mar. 2004), 45–47.

[12] D. Conte, P. Foggia, C. Sansone and M. Vento
Thirty years of graph matching in pattern recognition.
International Journal of Pattern Recognition & Artificial Intelligence, **18**, 3 (May 2004), 265–298.

[13] B. Das
Pepresenting uncertainties using Bayesian networks.
DSTO Electronics and Surveillance Research Laboratory (dsto-tr-0918), Salisbury South Australia, Australia, 199.

[14] N. Denis and E. Jones
Spatio-temporal pattern detection using dynamic Bayesian networks.
In *42nd IEEE Conference on Decision and Control*, vol. 5, 2003, 4533–4538.

[15] M.-L. Fernandez and G. Valiente
A graph distance metric combining maximum common subgraph and minimum common supergraph.
Pattern Recognition Letters, **22**, 6–7 (2001), 753.

[16] S. Fujita
Neighborhood information dissemination in the star graph.
IEEE Transactions on Computers, **49**, 12 (2000), 1366–1370.

[17] D. L. Hall and J. Llinas
An introduction to multisensor data fusion.
Proceedings of the IEEE, **85**, 1 (1997), 6.

[18] A. Hlaoui and W. Shengrui
A new algorithm for inexact graph matching.
In *Proceedings of 16th International Conference on Pattern Recognition*, vol. 4, 2002, 180–183.

[19] M. Ionescu and A. Ralescu
Fuzzy hamming distance as image similarity measure.
In *Proceedings of IPMU-2004*, Perugia, Italy, 2004, 1517–1523.

[20] M. Ionescu and A. Ralescu
Fuzzy hamming distance in a content-based image retrieval system.
In *IEEE International Conference on Fuzzy Systems*, vol. 3, 2004, 1721.

[21] M. Ionescu and A. Ralescu
The impact of image partition granularity using fuzzy hamming distance as image similarity measure.
In *Fifteenth Midwest Artificial Intelligence and Cognitive Sciences Conference*, Chicago IL, 2004, 111–118.

[22] M. Ionescu and A. Ralescu
Fuzzy hamming distance based banknote validator.
In *The 14th IEEE International Conference on Fuzzy Systems*, 2005, 300–305.

[23] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu
An efficient *k*-means clustering algorithm: analysis and implementation.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **24**, 7 (2002), 881.

[24] N. K. Kasabov
Hybrid fuzzy connectionist rule-based systems and the role of fuzzy rules extraction.
In *International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium*, vol. 1, 1995, 49–56.

[25] E. Kubicka, G. Kubicki, and I. Vakalis
Using graph distance in object recognition.
ACM Eighteenth Annual Computer Science Conference (CSC90), New York, NY, 1990, 43–48.

[26] N. C. Oza, A. N. Srivastava, and J. Stroeve
Improvements in virtual sensors: Using spatial information to estimate remote sensing spectra.
In *International Geoscience and Remote Sensing Symposium* (IGARSS), vol. 8, Seoul, South Korea, 2005, 5606–5609.

[27] C. Qu and Y. He
A method of threat assessment using multiple attribute decision making.
In *6th International Conference on Signal Processing*, vol. 2, 2002, 1091–1095.

[28] T. J. Ross
Fuzzy Logic with Engineering Applications.
New York: Wiley, 2nd ed., Aug. 1, 2004.

[29] P. J. Rousseeuw
Silhouettes: A graphical aid to the interpretation and validation of cluster analysis.
Journal of Computational and Applied Mathematics, 1987, 53–65.

[30] J. Salerno, M. Hinman, and D. Boulware
Building a framework for situation awareness.
Proceedings of the Seventh International Conference on Information Fusion, FUSION 2004, vol. 1, 2004, 219–226.

[31] K. Sambhoos, R. Nagi, M. Sudit and A. Stotz
Enhancements to high level data fusion using graph matching and state space search.
Submitted to Information Fusion, 2008.

[32] A. Schenker, M. Last, H. Bunke and A. Kandel
Classification of web documents using graph matching.
International Journal of Pattern Recognition and Artificial Intelligence, **18**, 3 (2004), 475–496.

[33] S. Selim and M. Ismail
K-means-type algorithms: A generalized convergence theorem and characterization of local optimality.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **6**, 1 (Jan. 1984), 81–87.

[34] H. Shiow-Fen and J. C. Gerard
k-neighborhood-covering and -independence problems for chordal graphs.
Society for Industrial and Applied Mathematics, **11**, 4 (1998), 633–643.

[35] A. Srivastava, N. Oza and J. Stroeve
Virtual sensors: Using data mining to efficiently estimate spectra.
IEEE Transactions on Geosciences and Remote Sensing, Special Issue on Advances in Techniques for Analysis of Remotely Sensed Data, **43**, 3 (2005), 590–600.

[36] A. Steinberg, C. Bowman and F. White
Revisions to the JDL data fusion model.
In *Proceedings of the SPIE—The International Society for Optical Engineering*, vol. 3719, 1999, 430–41.

[37] S. C. Stubberud and K. A. Kramer
Data association for multiple sensor types using fuzzy logic.
In *Proceedings of the IEEE Instrumentation and Measurement Technology Conference.*, vol. 3, 2005, 2154–2159.

[38] W. D. Wallis, P. Shoubridge, M. Kraetz and D. Ray
Graph distances using graph union.
Pattern Recognition Letters, **22**, 6–7 (2001), 701.

[39] W. Zhang
Depth-first branch-and-bound versus local search: A case study.
In *In Proceedings of the 17th National Conference on Artificial Intelligence*, 2000, 930–935.

Kedar Sambhoos received his B.E. degree in production engineering from the University of Pune, India, in 2002, and M.S. and Ph.D. degree from the Department of Industrial and Systems Engineering, State University of New York at Buffalo, NY, in 2004 and 2007, respectively.

He joined Information Fusion Division of CUBRC as a research scientist in February 2007. His current research interests are in general areas of information fusion with applications to soft and hard message data, and graph matching techniques applied to high level data fusion applications.

Rakesh Nagi is a Professor of Industrial and Systems Engineering at the University at Buffalo (SUNY). He received his Ph.D. (1991) and M.S. (1989) degrees in mechanical engineering from the University of Maryland at College Park, while he worked at the Institute for Systems Research and INRIA, France, and B.E. (1987) degree in mechanical engineering from the University of Roorkee (now IIT-R), India.

He is a recipient of Business First of Buffalo's "40 under Forty" award (2004), SME's Milton C. Shaw Outstanding Young Manufacturing Engineer Award (1999), IIE's Outstanding Young Industrial Engineer Award in Academia (1999), and National Science Foundation's CAREER Award (1996). His papers have been published in journals including *IIE Transactions*, *International Journal of Production Research*, *Journal of Manufacturing Systems*, *International Journal of Flexible Manufacturing Systems*, *Journal of Intelligent Manufacturing*, *Computers in Industry*, *Computer Integrated Manufacturing Systems*, *Operations Research*, *Naval Research Logistics*, *European Journal of Operational Research*, *Annals of Operations Research*, *Computers and Operations Research*, *Computers and Industrial Engineering*, and *ASME* and *IEEE Transactions*. Dr. Nagi's major research thrust is in the area of production systems and applied operations research. His research interests are in location theoretic approaches to facilities design, agile enterprises and information-based manufacturing, just-in-time production of assemblies, and information fusion.

Moises Sudit obtained his Bachelor of Science in industrial and systems engineering from Georgia Institute of Technology, his Master of Science in operations research from Stanford University and his Doctorate in operations research from Purdue University.

Dr. Sudit's primary research interests are in the theory and applications of discrete optimization. More specifically, he has been concerned in the design and analysis of methods to solve problems in the areas of integer programming and combinatorial optimization. One primary goal of this research has been the development of efficient exact and approximate (heuristic) procedures to solve large-scale engineering and management problems. As Managing Director of the Center for Multisource Information Fusion, Dr. Sudit has merged the interests of operations research with information fusion. He has an appointment as research professor in the School of Engineering and Applied Sciences at the University at Buffalo. He is a NRC Fellow through the Information Directorate at the Air Force Research Laboratory and has received a number of scholarly and teaching awards. He has a number of publications in distinguished journals and has been the principal investigator in numerous research projects.

John T. (Terry) Rickard (S'67—M'75—SM'01) received the B.S. and M.S. degrees in electrical engineering from Florida Institute of Technology, Melbourne, FL, in 1969 and 1971, respectively, and a Ph.D. degree in engineering physics from the University of California at San Diego, La Jolla, California, in 1975. He also received Series 7 and 63 General Securities Licenses in 1995 and a Series 24 General Securities Principal License in 1995.

He has 34 years of experience in technology and financial organizations, all of it in management and technology development positions. He began his career working in digital design and testing with Harris Corporation in 1969. He worked part-time as a graduate student for what was then the Naval Electronics Laboratory Center (now SPAWAR Systems Center) in San Diego, California from 1973 to 1975. In 1975, he cofounded ORINCON Corporation, a San Diego-based company specializing in the design and development of state-of-the-art data and information processing solutions for government and commercial customers. He ended his first career with ORINCON in 1994 as Senior Vice President and Technical Director. From 1994 to 2001, he served as President and later Chief Scientific Officer of OptiMark Technologies, Inc. He is a co-inventor of the OptiMark transaction matching system and was instrumental in the company's development from a start-up enterprise to an operating entity. Rejoining ORINCON in 2001 as senior vice president, his focus has been on broadening the company's technology base, particularly in machine intelligence. When ORINCON was acquired by Lockheed Martin in 2003, he was appointed to the position of Senior Principal Research Scientist. In 2005, he was elected a Senior Fellow of Lockheed Martin. In 2008, he retired from Lockheed Martin and joined Distributed Infinity, Inc., for whom he now works from his home in Larkspur, CO. His technical expertise includes signal processing, optimization, neural networks, fuzzy and expert systems, and graphical knowledge representation and inference for machine intelligence. His additional expertise includes financial engineering disciplines such as transaction systems, market structures, financial analytics, data mining, derivatives pricing, risk analysis, and trading strategies. His current research interests are in computational intelligence, conceptual spaces, information fusion, content based information retrieval, and nanotechnology.

Dr. Rickard has served on the boards of directors of three companies and has authored numerous technical publications that have appeared in refereed technical journals, books, and conference proceedings. In addition, he has authored several patents, and has several pending patent applications. He currently serves as the Vice Chairman of the IEEE Computational Intelligence Society, Denver Chapter and as a Technical Advisory Board member of a nanotechnology hedge fund. In 2006, he received the Author of the Year Award from Lockheed Martin Integrated Systems and Solutions.

Covariance Reconstruction for Track Fusion with Legacy Track Sources

YAAKOV BAR-SHALOM
HUIMIN CHEN

The problem of track-to-track association and track fusion has been considered in the literature where the fusion center has access to multiple track estimates and the associated estimation error covariances from local sensors, as well as their crosscovariances. Due primarily to the communication constraints in real systems, some legacy trackers may only provide the local track estimates to the fusion center without any covariance information. In some cases, the local (sensor-level) trackers operate with fixed filter gain and do not have any self assessment of their estimation errors. In other cases, the network conveys a coarsely quantized root mean square (RMS) estimation error of each local tracker. Thus the fusion center needs to solve the track association and fusion problem with incomplete data from legacy local trackers. The problem of track fusion with legacy track sources which lack covariances is handled by reconstructing them using sensor covariance and target maneuvering index information and then using the appropriate association and fusion algorithms. The situation when a coarsely quantized RMS estimation error is available is also discussed. A two-sensor tracking example is used to illustrate the effectiveness of the proposed covariance reconstruction method for track fusion and compared with a centralized interacting multiple model estimator.

Manuscript received February 9, 2007; revised August 14, 2008; released for publication October 23, 2008.

Refereeing of this contribution was handled by Benjamin Slocumb.

Authors' addresses: Yaakov Bar-Shalom, Univ. of Connecticut, Dept. of Electrical and Computer Engineering, Storrs, CT 06269-2157, USA, ybs@ee.uconn.edu; Huimin Chen, Univ. of New Orleans, Dept. of Electrical Engineering, New Orleans, LA 70148, USA, E-mail: (hchen2@uno.edu).

1. INTRODUCTION

In multisensor target tracking, each sensor can have its own target state estimate based on the local sensor measurements. Most existing communication networks between local trackers/sensors transmit to a fusion center the local track estimates—sometimes without any estimation error covariances, sometimes with partial covariance information and only rarely with full covariance information. In order to form a global picture of the existing tracks, it is necessary to associate multiple local tracks and fuse them to obtain the global target state estimates. Under this tracking configuration, the fusion center can carry out this association and fusion of the (latest) local track estimates on demand, which, in general, is less frequent than the measurement rate at each local sensor. Another important reason that track fusion (TrkF) is a viable alternative to centralized tracking (CenT), which requires transmission of all the measurements to the fusion center, is that the performance of TrkF is very close to that of CenT [4]. The problem of associating tracks represented by their local state estimates and covariances from multiple sources has been studied extensively in literature. While different sensors typically have independent measurement errors, the local state estimation errors for the same target are dependent due to the common process noise (and the prior, if common). This dependence is characterized by the crosscovariances of the local estimation errors [3]. Methods have been proposed to fuse the local tracks that carry out decorrelation [11, 12, 13]. Other techniques include track fusion that explicitly utilizes the crosscovariance information in a Bayesian setting [7, 10], with asynchronous sensors [1], and more generally, with possible common priors [15, 16, 17]. The work of [20] dealt with simultaneous general track-to-track association and bias estimation. In addition, the "covariance intersection" method proposed in [14] can fuse two estimates with unknown correlation. However, it is a very conservative method that avoids the issue of crosscovariances but may yield a fused covariance with diagonal elements that indicate a degradation in each component from the best estimate before fusion [9].

A legacy sensor and tracking system is one that was built in the past under different requirements, specifically, with no requirements to support network fusion. Thus no hardware/software facilities (or inadequate facilities) were included in the system to support the kind of track fusion that is desired now. To get the relevant data that one would like out of the system (i.e., covariances) requires a significant hardware/software modification to the system, which is impractical. Concisely, legacy can be defined as "you are stuck with what you've got." Before fusing local tracks, the fusion center has to decide whether they are from the same target. Track association is a hypothesis testing problem where local tracks are considered as having com-

mon origin (from the same target) vs. different ones by comparing a certain test statistic with a threshold to obtain desired test power [5, 18]. However, no previous results are available for the association and fusion of local tracks with legacy trackers that do not provide the necessary covariance information of the estimation errors.

In this paper we first consider the approximation of the covariance of the estimation error from a legacy tracker with a fixed filter gain. Then we use a two-sensor tracking scenario to compare the performance of the track fusion algorithm with the centralized target state estimator where the fusion center uses the state-of-the-art interacting multiple model (IMM) algorithm. Both the estimation accuracy and the credibility (consistency [2]) of the distributed tracker are compared with those of the centralized one. The results indicate that the performance degradation is small even during target maneuvers.

The rest of the paper is organized as follows. Section 2 describes the model used for legacy track sources. Section 3 presents a method to obtain the covariance of a legacy filter's track estimate as well as an approximation of the crosscovariance between two tracks. The reconstruction of the track covariance from a coarsely quantized estimation RMS error is also discussed. Section 4 presents a tracking example where two distributed tracking configurations are compared with the centralized estimator. Concluding remarks are provided in Section 5.

2. LEGACY TRACK SOURCES

In this section the model used for legacy track sources is formulated assuming the trackers are Kalman filters. To simplify the discussion, the model is presented for one generic coordinate with the target motion given by a discretized continuous time white noise acceleration (DCWNA) model [2]. For asynchronous sensors this model should be used to consistently handle the white process noise for all values of the sampling interval.

For sampling interval T, the state and measurement equations are

$$x(k+1) = Fx(k) + v(k) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x(k) + v(k) \quad (1)$$

$$z(k) = Hx(k) + w(k) = [1 \quad 0]x(k) + w(k) \quad (2)$$

where $v(k)$ is the zero mean white process noise sequence with covariance

$$E[v(k)v(k)'] \triangleq Q(t_{k+1} - t_k) = Q(T) = \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix} \tilde{q} \quad (3)$$

where $\tilde{q}$ is the (continuous time) process noise power spectral density (PSD)[1] and $w(k)$ is zero mean white measurement noise sequence, uncorrelated with the process noise, with variance

$$E[w(k)^2] = \sigma_w^2. \quad (4)$$

This describes the target motion along one dimension. For target motion in 2 or 3 dimensions, the model will consist of 2 or 3 such models with an appropriate stacked state vector.

The target maneuvering index, subscripted by "c" to indicate that it is based on the continuous time process noise [2], is defined as

$$\lambda_c = \sqrt{\frac{\tilde{q}T^3}{\sigma_w^2}}. \quad (5)$$

Then the steady state filter gain is

$$W = \begin{bmatrix} \alpha & \frac{\beta}{T} \end{bmatrix}' \quad (6)$$

where

$$\alpha = \beta\sqrt{u} \quad (7)$$

$$\beta = \frac{12}{6\left(u + \sqrt{u}\right) + 1} \quad (8)$$

$$u = \frac{1}{3} + \sqrt{\frac{1}{12} + \frac{4}{\lambda_c^2}}. \quad (9)$$

The steady state solution for the state estimation covariance matrix is given by

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix} = \begin{bmatrix} \alpha & \frac{\beta}{T} \\ \frac{\beta}{T} & \frac{\beta(\alpha - \beta/2)}{(1-\alpha)T^2} \end{bmatrix} \sigma_w^2. \quad (10)$$

The above solution is valid for the steady state of the DCWNA filter, but only with the *optimal* values of α and β as given in (7)–(8).

A legacy tracker uses a fixed gain W, not necessarily the optimal one, in each of its α-β filter updates and sends the state estimates to the fusion center, typically, without covariance information. Since track association and track fusion algorithms require such information in order to combine local tracks from different sources, a procedure to obtain this missing information is discussed next.

3. APPROXIMATION OF THE ESTIMATION ERROR COVARIANCE AND CROSSCOVARIANCE

Because of the time-varying target-sensor geometry, an α-β filter, even though it uses fixed gains, is not necessarily in steady state. This is due to the nonstationarity of the measurement noises, which is accounted for in

[1] See [2] on why it is incorrect to call this the variance of the process noise.

Subsection 3.1. Our model will assume that the tracking filter has a "slowly varying" (quasi-) steady state. The covariance of the target state estimate will be evaluated accounting for the fact that the sensor measurements (typically in polar or spherical coordinates for a radar), while having uncorrelated measurement noises between their components (range, azimuth/crossrange), have a coupling (correlation) between the track state estimation errors in different Cartesian coordinates. Subsection 3.2 deals with the case where the communication network provides partial covariance information in addition to the state estimates. A procedure to reconstruct the full state covariance matrix is presented. Since in the real world multiple sensors are practically never operating in a synchronized manner, the procedure for track fusion from asynchronous sensors is discussed in Subsection 3.3. Subsection 3.4 presents a simple method to approximate the crosscovariances between the state estimation errors of two local tracks from the same target by assuming constant correlation coefficients, whose exact values are shown to vary relatively little over the practical range of target maneuvering indices.

The hypothesis testing for track association and the fusion equations with the crosscovariance can be found in [3, Sec. 8.4].

3.1. Coupling Between Coordinates and Nonstationarity

For tracking in more than one dimension of the measurement space, the measurement covariance can be converted from the sensor coordinates (typically polar or spherical) into the coordinates in which the state is defined (usually Cartesian).[2] This will result in correlation between the state estimation errors in the Cartesian coordinates. It is important to preserve the coupling between the coordinates when the uncertainty ellipse for position is elongated and slanted, e.g., a "cigar" with the main axes at 45° and 135°. Neglecting the correlation between the coordinates would yield a much larger uncertainty region.

To preserve the coupling between the state space coordinates due to the measurements, the fusion center should run the Joseph form of the covariance update iteration[3] at time k [2]

$$P(k) = [I - WH][FP(k-1)F' + Q][I - WH]' + WR(k)W' \tag{11}$$

with the appropriate sampling interval. The Joseph form is needed because the legacy filter gain is not optimal and only this equation is valid for the covariance (actually MSE matrix) update when *arbitrary* filter gains are used. The process noise covariance Q should be selected by the fusion center to model the target motion uncertainty to the extent possible. The filter gain W in (11) should be the same as in the legacy filter. If W is not known at the fusion center, it should be "replicated" using (6). The measurement noise covariance $R(k)$, assumed to be known,[4] in (11) is the covariance of the measurements converted from polar to Cartesian. The measurement conversion should be linearized at the latest measurement or the measurement prediction using the latest state. When $P(k-1)$ is unavailable at the fusion center, one can assume that $P(k-1) = P(k)$ in (11), resulting in an algebraic Riccati equation. This will yield a (slowly) time-varying covariance matrix that accounts for the nonstationarity of measurement noise.

3.2. Approximation of the Estimation Error Covariance of Legacy Trackers with Partial Information

When the communication network can provide a coarsely quantized (i.e., an approximate) 2-dimensional root mean square (RMS) position error, denoted as RMS_p, to the fusion center, the state estimation error covariance can be obtained as follows.

We shall model RMS_p as the (steady state) error of two independent α-β filters, one in the range direction, the other in the cross-range direction. Denote the measurement noise RMS values in these directions as σ_r and $\sigma_{\times r}$, respectively. These are assumed to be known, based on the radar specifications and the radar-target geometry.

The position gains for these two filters are, according to (6),

$$\alpha_r = \alpha(\lambda_{c_r}) \tag{12}$$

and

$$\alpha_{\times r} = \alpha(\lambda_{c_{\times r}}) \tag{13}$$

respectively, where the corresponding target maneuvering indices are, similarly to (5),

$$\lambda_{c_r} = \sqrt{\frac{\tilde{q}T^3}{\sigma_r^2}} \tag{14}$$

and

$$\lambda_{c_{\times r}} = \sqrt{\frac{\tilde{q}T^3}{\sigma_{\times r}^2}} \tag{15}$$

with $\tilde{q}$ the continuous time process noise PSD that models the motion uncertainty (in both the range and cross-range directions, uncorrelated between them) and T the sampling interval.

The RMS position error from the above filters is, based on (10), given by

$$\text{RMS}_p = \sqrt{\alpha_r \sigma_r^2 + \alpha_{\times r} \sigma_{\times r}^2}. \tag{16}$$

[2]While some tracking systems keep the measurements in polar/spherical coordinates, the conversion to Cartesian allows exact debiasing when necessary [2].

[3]A single time argument is used here for the covariance.

[4]The measurement noise variances are, particularly in azimuth/elevation (and thus in cross-range), dependent on the target SNR (inversely proportional to the SNR [6]; in range the variance depends primarily on the pulse waveform). However, unless one assumes these variances as known (for an "average" SNR), one cannot reconstruct the track errors. Consequently, $R(k)$ is assumed known.

Assuming the value of RMS_p is available and the measurement noise variances are known (as discussed previously), one can solve (16) (after substituting (14)–(15) into (12)–(13) and the result into (16)) to find $\tilde{q}$. Once this is obtained, one can use (10) or (11) to reconstruct (approximately) the covariance of the entire state estimate. Note, however, that while this is in a Cartesian coordinate system, this system is aligned with the line of sight from the radar to the target and it has to be rotated into the local (common) Cartesian system, which is, typically, East-North.

The above procedure allows to reconstruct (approximately) the estimation error covariance of the legacy tracker from a coarsely quantized position RMS error, which is assumed to be conveyed by a communication network. A similar approach can be taken when RMS_p is a position prediction error, as well as for the 3-dimensional case.

3.3. Prediction to Fusion Time for Asynchronous Sensors

For asynchronous sensors, the state prediction (to the time for which fusion will be carried out) based on the legacy tracker's latest estimate should be used by the fusion center. Assume that the fusion is done at time k and the most recent estimate at the fusion center from the legacy tracker is $\hat{x}(\kappa)$ at time[5] κ, with $\kappa < k$. Then the fusion center needs to (i) approximate the estimation error covariance[6] $P(\kappa)$ at time κ using (10) or (11) and (ii) apply the standard prediction equations given by

$$\hat{x}(k) = F(k,\kappa)\hat{x}(\kappa) \tag{17}$$

$$P(k) = F(k,\kappa)P(\kappa)F(k,\kappa)' + Q(k,\kappa) \tag{18}$$

to obtain the state prediction and the corresponding error covariance for time k. For the motion model (1), $Q(k,\kappa)$ is given by (3) with $T = t_k - t_\kappa$.

Thus what is needed to evaluate the covariance of the estimate from a legacy tracker are:

- the sampling times
- the process noise PSD
- the measurement noise covariance.

It should be noted that the parameters based on which the legacy tracker has been designed are unlikely to be the same as listed above. Thus, what the fusion center should do is to replicate the performance of the legacy tracker to the extent possible.

[5]We use for simplicity the notations κ and k instead of t_κ and t_k.

[6]A single time argument is used here for the covariance. This covariance can be an updated covariance at the current time for one sensor, or a prediction to the current time for another sensor.

3.4. Approximation of the Crosscovariance of the Estimation Errors

When two local tracks have correlated estimation errors, assuming they are operating synchronously[7] and use the same target motion and measurement models, in the steady state, the crosscovariance matrix is given by [3]

$$P^\times = [I - WH][FP^\times F' + Q][I - WH]'. \tag{19}$$

The above Lyapunov type matrix equation can be solved numerically for any given target maneuvering index by simple forward iteration starting from $P^\times = 0$. For a distributed tracking system, the calculation of the crosscovariance using (19) is not practical.

The following approximation is considered [8]. Denote by P^{ij} the approximate crosscovariance matrix between local tracks i and j. Each element of P^{ij}, which is a 2×2 matrix for the model considered in (1), is approximated by constant correlation coefficients as follows

$$P^{ij}_{lm} = \rho_{lm}[P^i_{ll}P^j_{mm}]^{1/2}, \qquad l,m = 1,2 \tag{20}$$

where ρ_{11} is the position-position correlation coefficient, ρ_{12} is the position-velocity correlation coefficient and ρ_{22} is the velocity-velocity correlation coefficient.

Assuming equal variances of the measurement error for both sensors, we can solve the Lyapunov equation for the steady state DCWNA model. The resulting crosscorrelation coefficients between the estimation errors from the two local trackers, namely, ρ_{11}, ρ_{22} and ρ_{12}, are shown in Fig. 1 for target maneuvering index values within $[0.05, 2]$. These results are similar to those in [8] where the discrete time white noise acceleration model (DWNA) [2] is used. For the simulations to be presented in Section 4, we choose, in view of the fact that, as it can be seen from Fig. 1, these coefficients are nearly constant, the following fixed values

$$\rho_{11} = 0.15, \qquad \rho_{12} = 0.25, \qquad \rho_{22} = 0.7 \tag{21}$$

to compute the approximate crosscovariance according to (20).

Legacy trackers can be assumed as being decoupled across coordinates since the process noise is assumed uncorrelated between different coordinates. Consequently, the crosscorrelation between one sensor's tracking errors in one coordinate and another sensor's tracking errors in another coordinate will be zero due to the lack of common process noise. Thus the crosscovariance matrix will be assumed to have blocks consisting of zeros between different coordinates.

[7]The exact general recursion of the crosscovariance for asynchronous sensors is presented in the Appendix.

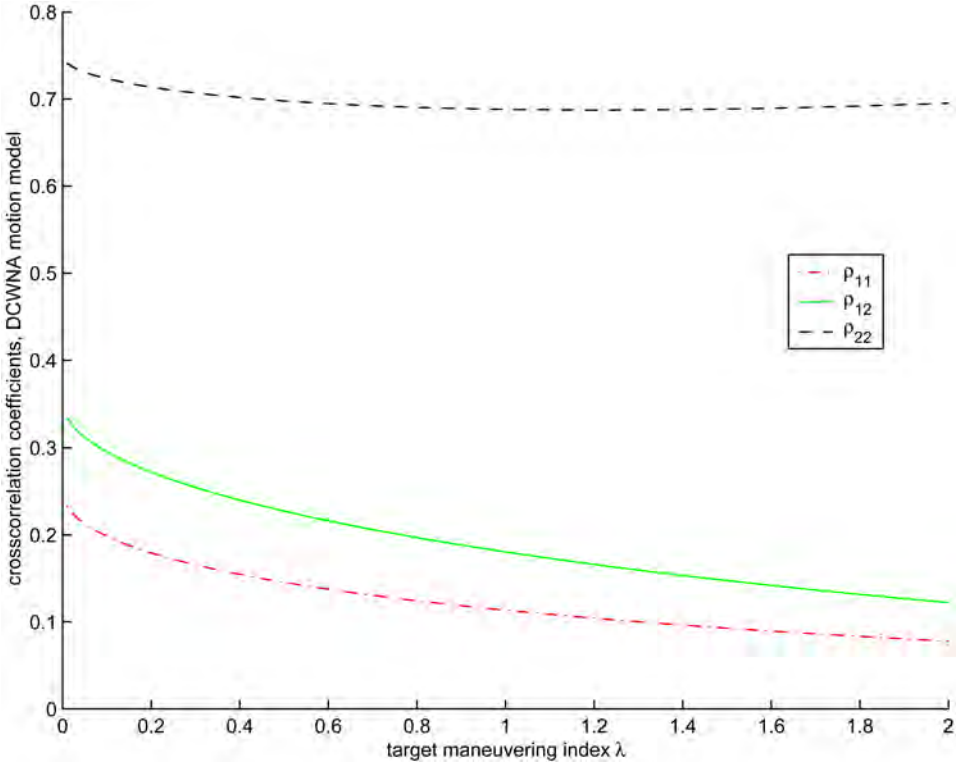


Fig. 1. Crosscorrelation coefficients vs. target maneuvering index for DCWNA model.

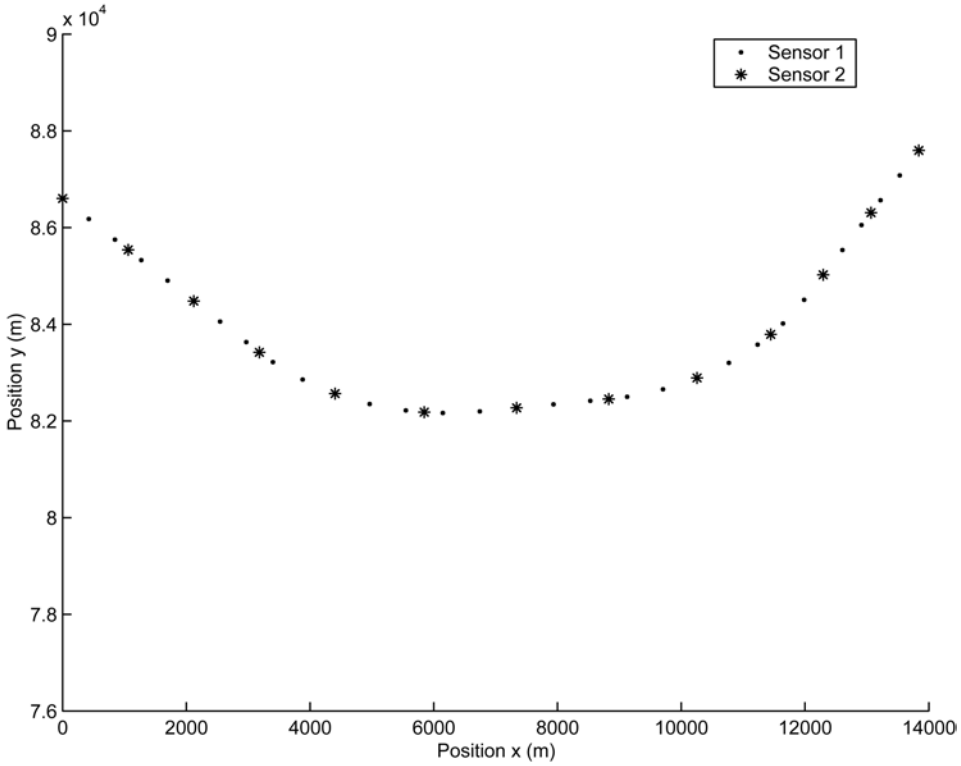


Fig. 2. Target trajectory with true positions where measurements are made by the two sensors.

4. EXAMPLE OF TRACK FUSION WITH A LEGACY TRACK SOURCE

We consider a ground target tracking scenario where two sensors are located at $(-50,0)$ km and $(50,0)$ km, respectively. Both sensors measure the target range and bearing with the same standard deviations of the measurement error given by $\sigma_r = 50$ m and $\sigma_b = 2$ mrad. The sampling interval of sensor 1 is $T_1 = 2$ s while the sampling interval of sensor 2 is $T_2 = 5$ s.

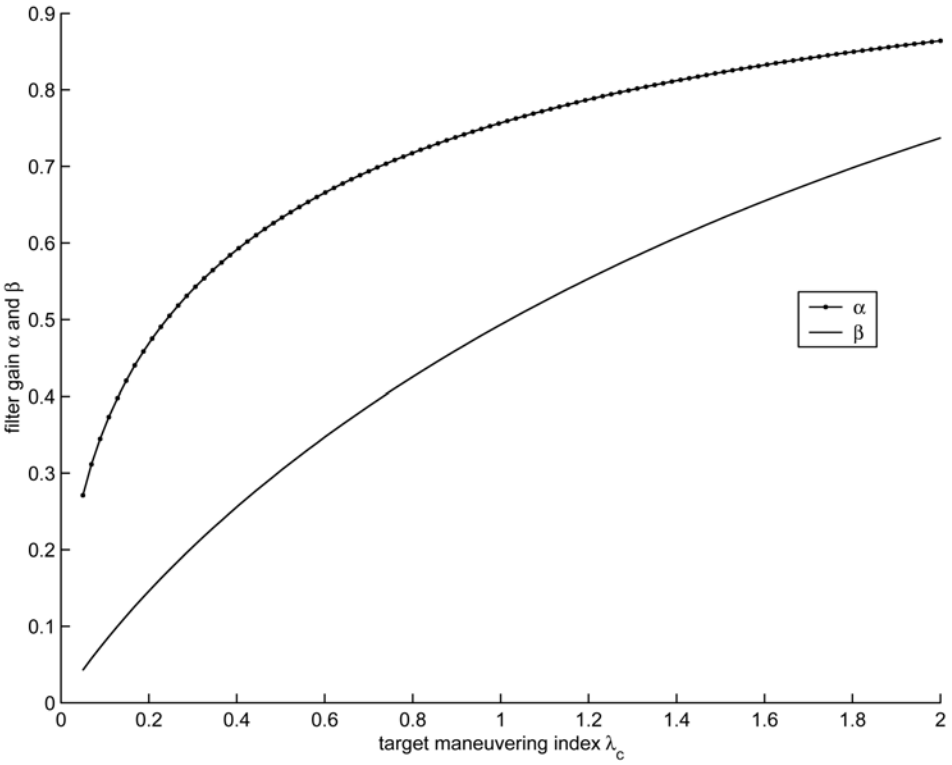


Fig. 3. Steady state filter gains vs. target maneuvering index for DCWMA model.

The target is initially at (0,86.6) km moving at a speed of 300 m/s toward south-east on a course of approximately $-135°$. Then at $t = 15$ s the target makes a course change with a constant turn rate of 4°/s (acceleration of about 2.1 g over a duration T_{man} of about 11 s) and heads toward east. The target makes a second course change at $t = 35$ s with a constant turn rate of 4°/s and heads toward north-east. The target trajectory is shown in Fig. 2 where the true target positions are indicated at the time instances at which a measurement is made by sensor 1 or sensor 2. The total time for the target to complete the designated trajectory is 60 s. Note that the target range is around 100 km at the beginning for both sensors, where the standard deviation of the crossrange measurement error is around 200 m. The true target motion has no process noise in this case.

We consider the following three tracking configurations for performance comparison.

(i) A centralized estimator which uses an IMM with two models and sequentially updates the target state with measurements from both sensors. This IMM estimator has a DCWNA model with low process noise PSD $\tilde{q}_l$ to capture the uniform target motion and a DCWNA model with high process noise PSD $\tilde{q}_h$ to capture the two turns. We use $\tilde{q}_l = 1$ m^2/s^3 and $\tilde{q}_h = 8000$ m^2/s^3 which, for $T_1 = 2$ s, corresponds to a target maneuvering of $\sqrt{\tilde{q}_h/T_1} \approx 6.4$ g. The process noise is the same in east and north of the Cartesian coordinates and uncorrelated between these coordinates. The transition between the modes is modeled according to a continuous time Markov chain with the expected sojourn times [2, Sec. 11.7.3] in these modes given by $1/\lambda_1$ and $1/\lambda_2$, respectively. These correspond to exponential sojourn time distributions with parameters λ_1 and λ_2, respectively. The transition probability matrix between the two models (generalized version of Eq. (11.6.7-1) in [2]) from any time t_1 to time t_2 is [19]

$$\Pi(t_2,t_1) = \frac{1}{\lambda_1+\lambda_2}\begin{bmatrix} \lambda_2+\lambda_1 e^{-(\lambda_1+\lambda_2)T} & \lambda_1-\lambda_1 e^{-(\lambda_1+\lambda_2)T} \\ \lambda_2-\lambda_2 e^{-(\lambda_1+\lambda_2)T} & \lambda_1+\lambda_2 e^{-(\lambda_1+\lambda_2)T} \end{bmatrix} \tag{22}$$

where $T = |t_2 - t_1|$. For the scenario used in simulation, we chose $\lambda_1 = (1/20)$ s^{-1} and $\lambda_2 = (1/10)$ s^{-1}.

(ii) In the first decentralized tracking configuration both sensor 1 and sensor 2 use an IMM estimator and the fusion center fuses the local estimates every $T_F = 10$ s using the two local state estimates with the corresponding covariances. The local tracker at sensor 1 uses $\tilde{q}_l = 1$ m^2/s^3 and $\tilde{q}_h = 8000$ m^2/s^3. The local tracker at sensor 2 uses $\tilde{q}_l = 1$ m^2/s^3 and $\tilde{q}_h = 20000$ m^2/s^3.

(iii) In the second decentralized tracking configuration sensor 1 uses the same IMM estimator as in (ii) while sensor 2 uses a legacy filter (in both north and east coordinates) with a fixed filter gain. The optimal filter gain components α and β vs. the target maneuvering index are shown in Fig. 3. The values we used are $\alpha = 0.86$ and $\beta = 0.74$ which correspond to a target maneuvering index around 2. To implement the track-to-track fusion with a legacy tracker, the fusion center needs first to obtain the covariance (approximate MSE matrix) of the local state estimate from the legacy tracker. This is done according to the procedure discussed in Section 3.1. The track fusion is done with

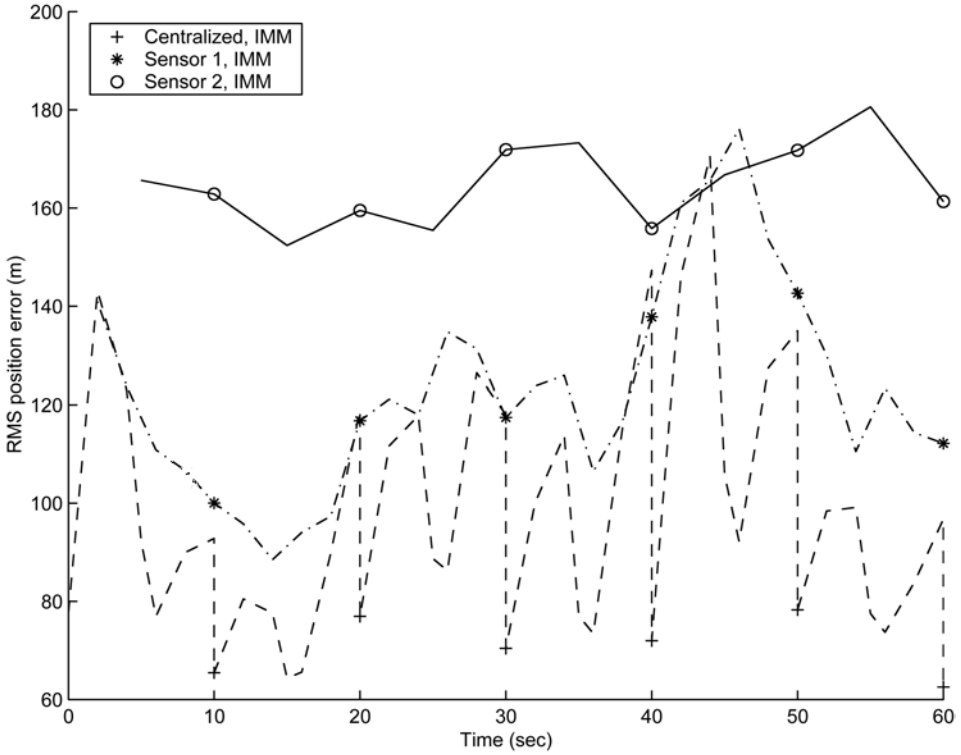


Fig. 4. RMS position errors for centralized IMM estimator vs. two local IMM estimators.

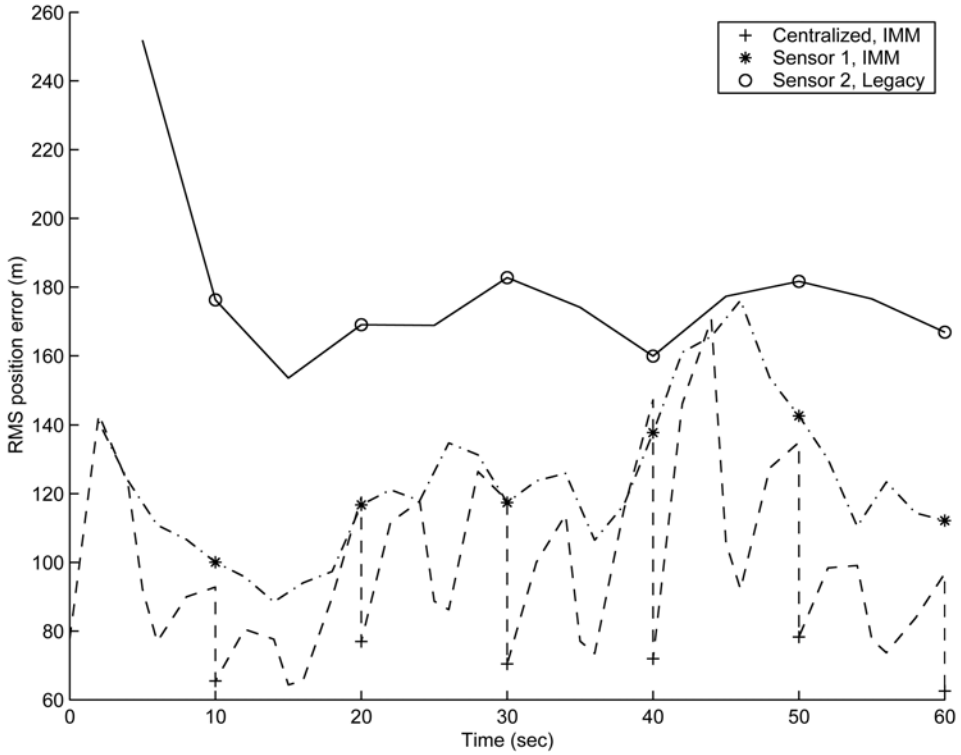


Fig. 5. RMS position errors for centralized IMM estimator vs. two local estimators (IMM and legacy).

crosscovariances calculated using the fixed crosscorrelation coefficients as in (ii).

Fig. 4 shows the root mean square (RMS) position errors of the centralized IMM estimator vs. the two local IMM estimators at sensor 1 and sensor 2 from 100 Monte Carlo runs. Special symbols indicate the times when track-to-track fusion is carried out. The local tracker at sensor 1 has better estimation accuracy in position than the local tracker at sensor 2 since it has a higher measurement rate.

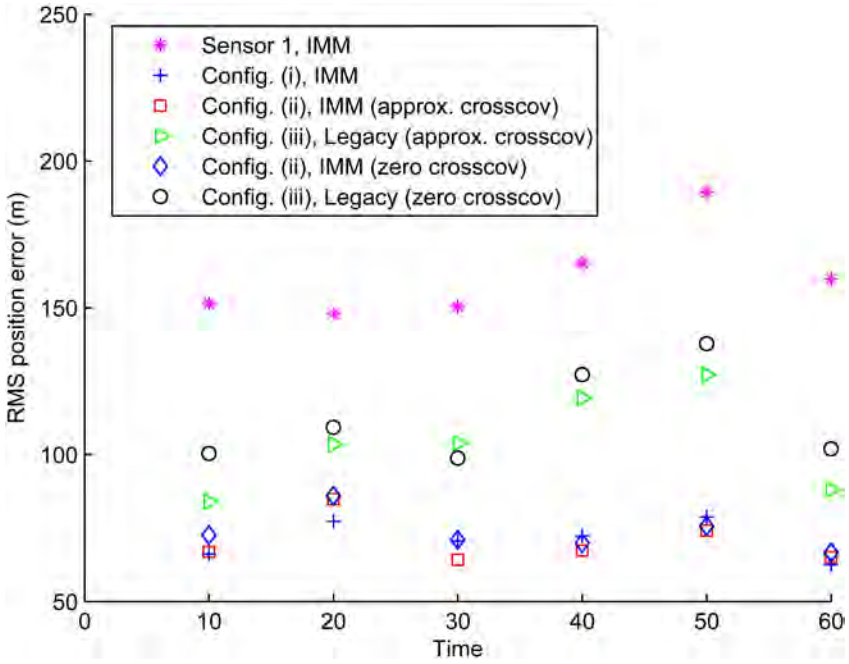


Fig. 6. Comparison of the RMS position errors for centralized IMM estimator vs. track fusion with an IMM estimator and a legacy filter.

Fig. 5 shows the RMS position errors of the centralized IMM estimator vs. two local estimators at sensor 1 and sensor 2 where the local tracker at sensor 1 uses an IMM estimator and the local tracker at sensor 2 uses a legacy filter from 100 Monte Carlo runs. Compared with Fig. 4, we can see that the performance degrades when sensor 2 uses a legacy filter rather than an IMM estimator.

Fig. 6 shows the RMS position errors at the fusion center for the above three tracking configurations as well as that by sensor 1 alone. In configurations (ii) and (iii), both approximate crosscovariance and zero crosscovariance were used in the track fusion procedure. We can see that the track fusion of two local IMM estimators has the RMS position error close to that of the centralized estimator. Assuming zero crosscovariance does not affect the position estimation accuracy by much. However, the track fusion with a legacy filter has a moderate performance gap compared with the centralized estimator for the RMS position error. We can also see that the performance of the fused estimate using a legacy track is clearly better than that using sensor 1 alone.

Fig. 7 shows the normalized estimation error squared (NEES) [2] at the fusion center for the above three tracking configurations as well as that by sensor 1 alone. The 99% percent confidence interval is also shown assuming that the NEES statistic is chi-square distributed with the appropriate degrees of freedom. We can see that nearly all fusion results are pessimistic during the non-maneuver motion segments owing to the zero process noise of the true target motion. However, configurations (ii) and (iii) yield larger NEES than the centralized estimator during the target turns, i.e., the estimation error covariance at the fusion center is more optimistic compared with that of the single sensor estimate. Assuming zero crosscovariance can make the situation even worse. Thus caution has to be exercised when fusing local estimates that are not consistent with their calculated covariances.

5. CONCLUDING REMARKS

In this paper a procedure for reconstruction of legacy trackers' state estimation error covariances was described for use in track-to-track association and fusion algorithms that account for the crosscovariance of the estimation errors between local tracks. In addition, a practical way to approximate these crosscovariances has been presented. A two-sensor tracking example, with one of the trackers being a legacy tracker, indicates the effectiveness of the resulting distributed tracking system with track fusion on demand. The performance of this system exhibits only a modest degradation compared with a centralized tracker using an interacting multiple model estimator.

APPENDIX. THE EXACT CROSSCOVARIANCE FOR ASYNCHRONOUS SENSORS

The recursion for the crosscovariance given in Eq. (8.4.2-3) of [3] is for synchronous sensors. The recursion for the case of asynchronous sensors is as follows.

Let $\{t_m^i\}_{m=1}^{N^i}$ and $\{t_n^j\}_{n=1}^{N^j}$ be the sampling times at sensor i and j, respectively. The union of these sets, with the times ordered, is denoted as

$$\mathcal{T}^{ij} \triangleq \{t_k^{ij}\}_{k=1}^{N^i+N^j}. \tag{23}$$

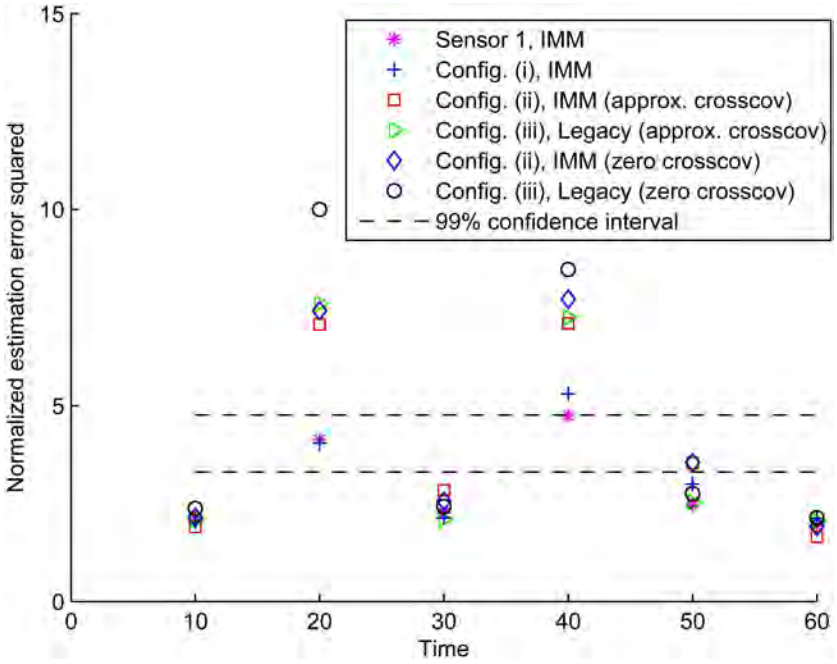


Fig. 7. Comparison of the NEES for centralized IMM estimator vs. track fusion with an IMM estimator and a legacy filter.

Then the generalized version of the crosscovariance recursion will be iterated on the ordered union set (23) as follows

$$P^{ij}(t_k^{ij}) \triangleq E[\tilde{x}^i(t_k^{ij})\tilde{x}^j(t_k^{ij})']$$
$$= [I - W^i(t_k^{ij})H^i(t_k^{ij})][F(t_k^{ij},t_{k-1}^{ij})P^{ij}(t_k^{ij}-1)F(t_k^{ij},t_{k-1}^{ij})'$$
$$+ Q(t_k^{ij},t_{k-1}^{ij})][I - W^j(t_k^{ij})H^j(t_k^{ij})]' \quad (24)$$

where the estimation error $\tilde{x}$ has t_k^{ij} as its single argument indicating the current time. This error might correspond to a current estimate, or a prediction as in (17). The gain for filter i in the above is

$$W^i(t_k^{ij}) = 0 \qquad \text{if} \quad t_k^{ij} = t_n^j \quad (25)$$

i.e., it is zero at the times when only filter j carries out an update, and the other way around; $F(t_k^{ij},t_{k-1}^{ij})$ is the state transition matrix from t_{k-1}^{ij} to t_k^{ij} and $Q(t_k^{ij},t_{k-1}^{ij})$ is the covariance of the process noise over the interval $t_k^{ij} - t_{k-1}^{ij}$. The initial condition for (25) is $P^{ij}(t_1^{ij}) = 0$, assuming the filters use independent initial information.

REFERENCES

[1] A. T. Alouani and J. E. Gray
Theory of Distributed Estimation Using Multiple Asynchronous Sensors.
IEEE Transactions on Aerospace and Electronic Systems, **41**, 2 (Apr. 2005), 717–722.

[2] Y. Bar-Shalom, X. R. Li and T. Kirubarajan
Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software.
New York: Wiley, 2001.

[3] Y. Bar-Shalom and X. R. Li
Multisensor, Multitarget Tracking: Principles and Techniques.
Storrs, CT: YBS Publishing, 1995.

[4] Y. Bar-Shalom
Hierarchical tracking for the real world.
IEEE Transactions on Aerospace and Electronic Systems, **42**, 3 (July 2006), 846–850.

[5] Y. Bar-Shalom and H. Chen
Multisensor track-to-track association for tracks with dependent errors.
Journal of Advances in Information Fusion, **1**, 1 (July 2006), 3–15.

[6] W. D. Blair, G. A. Watson, T. Kirubarajan and Y. Bar-Shalom
Benchmark for radar allocation and tracking targets in ECM.
IEEE Transactions on Aerospace and Electronic Systems, **34**, 4 (Oct. 1998), 1097–1114.

[7] K. C. Chang, Z. Tian, S. Mori and C. Y. Chong
MAP track fusion performance evaluation.
In *Proceedings of the 2002 International Conference on Information Fusion*, Annapolis, MD, USA, July 2002.

[8] H. Chen, T. Kirubarajan and Y. Bar-Shalom
Performance limits of track-to-track fusion vs. centralized estimation: Theory and application.
IEEE Transactions on Aerospace and Electronic Systems, **39**, 2 (Apr. 2003), 386–400.

[9] H. Chen and Y. Bar-Shalom
Track fusion with legacy track sources.
In *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, July 2006.

[10] C. Y. Chong, S. Mori, W. H. Barker and K. C. Chang
Architectures and algorithms for track association and fusion.
IEEE Aerospace and Electronic Systems Magazine, Jan. 2000, 5–13.

[11] O. E. Drummond
Track fusion with feedback.
In *Proceedings of the 1996 SPIE Conference on Signal and Data Processing of Small Targets*, vol. 2759, 1996, 342–360.

[12] O. E. Drummond
A hybrid sensor fusion algorithm architecture and tracklets.
In *Proceedings of the 1997 SPIE Conference on Signal and Data Processing of Small Targets*, vol. 3163, 1997, 512–524.

[13] O. E. Drummond, W. D. Blair, G. C. Brown, T. L. Ogle, Y. Bar-Shalom, R. L. Cooperman and W. H. Barker
Performance assessment and comparison of various tracklet methods for maneuvering targets.
In *Proceedings of the SPIE Conference on Signal Processing, Sensor Fusion, and Target Recognition XII*, vol. 5096, Apr. 2003.

[14] S. Julier and J. K. Uhlmann
General decentralized data fusion with covariance intersection.
In D. L. Hall and J. Llinas (eds.) *Handbook of Multisensor Data Fusion*, CRC Press, 2001, ch. 12.

[15] X. R. Li
Applied Estimation and Filtering.
Lecture Notes, University of New Orleans, preprint, 2002.

[16] X. R. Li, Y. Zhu, J. Wang and C. Han
Optimal linear estimation fusion, Part I: Unified fusion rules.
IEEE Transactions on Information Theory, **49**, 9 (Sept. 2003), 2192–2208.

[17] X. R. Li
Optimal linear estimation fusion, Part VII: Dynamic systems.
In *Proceedings of the 2003 International Conference on Information Fusion*, Cairns, Australia, July 2003.

[18] X. D. Lin, Y. Bar-Shalom and T. Kirubarajan
Multisensor bias estimation using local tracks without a priori association.
In *Proceedings of the SPIE Conference on Signal and Data Processing of Small Targets*, vol. 5204, San Diego, CA, Aug. 2003.

[19] A. Papoulis and S. U. Pillai
Probability, Random Variables and Stochastic Processes.
McGraw Hill, 2002.

[20] L. D. Stone, M. Williams and T. Tran
Track-to-track association and bias removal.
In *Proceedings of the SPIE Conference on Signal and Data Processing of Small Targets*, vol. 4728, Orlando, FL, Apr. 2002.

Yaakov Bar-Shalom was born on May 11, 1941. He received the B.S. and M.S. degrees from the Technion, Israel Institute of Technology, in 1963 and 1967 and the Ph.D. degree from Princeton University in 1970, all in electrical engineering.

From 1970 to 1976 he was with Systems Control, Inc., Palo Alto, CA. Currently he is Board of Trustees Distinguished Professor in the Dept. of Electrical and Computer Engineering and Marianne E. Klewin Professor in Engineering at the University of Connecticut. He is also Director of the ESP (Estimation and Signal Processing) Lab.

His current research interests are in estimation theory and target tracking and has published over 370 papers and book chapters in these areas and in stochastic adaptive control. He coauthored the monograph *Tracking and Data Association* (Academic Press, 1988), the graduate texts *Estimation and Tracking: Principles, Techniques and Software* (Artech House, 1993), *Estimation with Applications to Tracking and Navigation: Algorithms and Software for Information Extraction* (Wiley, 2001), the advanced graduate text *Multitarget-Multisensor Tracking: Principles and Techniques* (YBS Publishing, 1995), and edited the books *Multitarget-Multisensor Tracking: Applications and Advances* (Artech House, Vol. I, 1990; Vol. II, 1992; Vol. III, 2000).

He has been elected Fellow of IEEE for "contributions to the theory of stochastic systems and of multitarget tracking." He has been consulting to numerous companies and government agencies, and originated the series of Multitarget-Multisensor Tracking short courses offered via UCLA Extension, at Government Laboratories, private companies and overseas.

During 1976 and 1977 he served as Associate Editor of the IEEE Transactions on Automatic Control and from 1978 to 1981 as Associate Editor of Automatica. He was Program Chairman of the 1982 American Control Conference, General Chairman of the 1985 ACC, and Co-Chairman of the 1989 IEEE International Conference on Control and Applications. During 1983–87 he served as Chairman of the Conference Activities Board of the IEEE Control Systems Society and during 1987–89 was a member of the Board of Governors of the IEEE CSS. He was a member of the Board of Directors of the International Society of Information Fusion (1999–2004) and served as General Chairman of FUSION 2000, President of ISIF in 2000 and 2002 and Vice President for Publications in 2004–08.

In 1987 he received the IEEE CSS Distinguished Member Award. Since 1995 he is a Distinguished Lecturer of the IEEE AESS and has given numerous keynote addresses at major national and international conferences. He is corecipient of the M. Barry Carlton Award for the best paper in the IEEE Transactions on Aerospace and Electronic Systems in 1995 and 2000 and the 1998 University of Connecticut AAUP Excellence Award for Research. In 2002 he received the J. Mignona Data Fusion Award from the DoD JDL Data Fusion Group. He is a member of the Connecticut Academy of Science and Engineering. He is the recipient of the 2008 IEEE Dennis J. Picard Medal for Radar Technologies and Applications.

Huimin Chen received the B.E. and M.E. degrees from Department of Automation, Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, in 2002, all in electrical engineering.

He was a post doctorate research associate at Physics and Astronomy Department, University of California, Los Angeles, and a visiting researcher with the Department of Electrical and Computer Engineering, Carnegie Mellon University from July 2002 where his research focus was on weak signal detection for single electron spin microscopy. He joined the Department of Electrical Engineering, University of New Orleans in Jan. 2003 as an assistant professor. His research interests are in general areas of signal processing, estimation theory, and information theory with applications to target detection and target tracking.

Intelligent System for Interpreting the Pattern of Volcanic Eruptions

GALINA L. ROGOVA
MARCUS I. BURSIK
SARA HANSON-HEDGECOCK

The overall goal of the research presented in this paper is to design an intelligent system to aid geologists in processing complex geologic characteristics for interpreting eruption patterns, and thereby to aid eruption forecasting for volcanic chains and fields. The objective of this paper is twofold. First it describes applications of data fusion techniques to designing such an intelligent system. The paper discusses the system architecture and applicability and benefits of evidential decision fusion methods for processing uncertain rock characteristics. Second, it introduces a new evidential method of combining several clustering results and presents the results of application of this fusion method to clustering geochemical data characterizing volcano magma chambers.

Manuscript received January 6, 2008; revised November 4, 2008; released for publication December 2, 2008.

Refereeing of this contribution was handled by Jean Dezert.

This research was supported by the University at Buffalo, and by the NSF under grant EAR0538227.

Authors' addresses: G. Rogova, Encompass Consulting, 9 Country Meadows Drive, Honeoye Falls, NY 14472, E-mail: (rogova @rochester.rr.com); M. Bursik, Geology Dept., State University of New York at Buffalo, Buffalo, NY 14260, E-mail: (mib@geology.buffalo.edu); S. Hanson-Hedgecock, Geology Dept., State University of New York at Buffalo, Buffalo, NY 14260, E-mail: (seh5 @buffalo.edu).

1557-6418/08/$17.00

1. INTRODUCTION

Volcanoes erupt mixtures of gas and rocks (generically known as tephra) [37]. The tephra settles to the earth's surface and leaves a record of the eruption. By looking at the separate tephra layers preserved within the soil layers, we are able to understand the history of eruptions of a volcano. Because volcanoes are creatures of habit, they tend to act in the future as they did in the past. Thus we are able to forecast future behaviour by observing the features of the tephra layers from past eruptions. Unfortunately, the preservation of the tephra layers is not complete. Erosion removes the tephra from many locations, and eventually the tephra is buried under enough younger layers that it is difficult to reach by excavation. Thus the data may be sparse. Variability within the tephra grains, and insufficient sampling often result in relatively large variances and imprecision in the dataset. Finally, we must match (correlate) the same tephra layer from one locality to another to characterize the layer thoroughly and understand its story. However, the correlation process is rarely straightforward owing to uncertainties about specific tephra layer identity.

There are two groups of characteristics that are used for the correlation process: physical and geochemical. Physical features include such variables as layer thickness, size of grains of different types, arrangement of the grains within the layer, and relative abundance of the different grain types (Fig. 1).

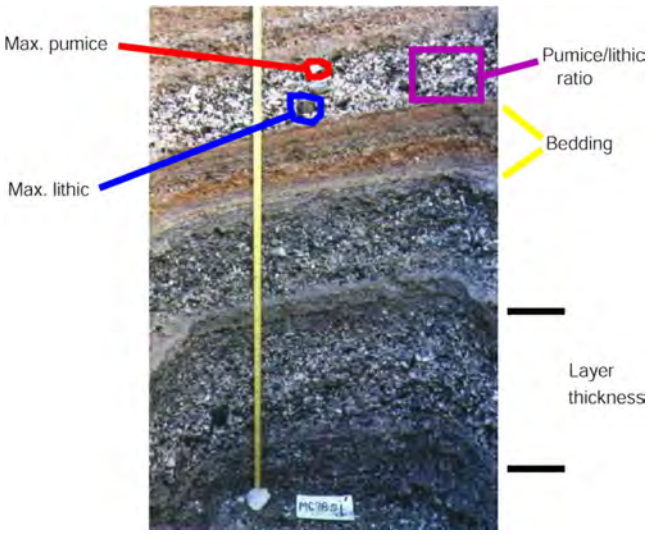


Fig. 1. Physical features of tephra layers shown in a hand-dug pit. The white rocks are pumice; black rocks are mostly obsidian.

Some of the physical characteristics (e.g., thikness) are easy to observe and measure while "in the field" and looking at the layers; others represent expert opinion on their values. Thus, the physical characteristics are the primary features used in identification and correlation. However recognition of stratigraphic patterns among different workers, and even for one worker, can be difficult because of the variability. The result of such

analysis depends on one's level of expertise and can be subjective and time consuming. Therefore it is beneficial to develop an intelligent system to aid geologists in recognition of tephra layers.

The geochemical make-up of a tephra layer helps us not only with correlation, but also tells us much about the reservoir (magma chamber) from which the tephra layer was erupted. A magma chamber is a subterrenean feature and is therefore inaccessible to direct observation. The chemical make-up of the rocks that came from the chamber therefore carries indicators of some features of the chamber. For example, by observing the concentrations of different elements found within different rocks, we can determine whether there was one coherent batch of magma responsible for the layers, or whether there were different, but variously interconnected batches.

The paper presents the first attempt to develop a systematic approach to processing complex geologic characteristics for interpreting eruption patterns. This processing utilizes decision fusion techniques developed in the framework of the Transferable Belief Model [35, 36]. Although utilization of different types of data by geologists has been used in the study of seismic and other geophysical data relating to volcanic eruption forecasting [5], application of information fusion techniques in eruption forecasting is a new field [8, 31].

The paper comprises two parts. The first one describes an intelligent system for assisting geologists in processing uncertain geological data for interpreting eruption patterns as well as the applicability and benefits of evidential decision fusion methods for processing uncertain rock characteristics. The second part introduces an evidential method of combining several clustering results and application of this method to defining magma chambers by clustering vent geochemical characteristics.

2. DECISION FUSION FOR INTERPRETING THE PATTERN OF VOLCANO ERUPTION

2.1. The Processing

Information flow in the intelligent system of stratigraphic layer characterization designed to support geologists in interpretation of eruption patterns loosely follows the major steps of geological data analysis performed by geologists (Fig. 2), which is comprised of two interrelated tasks:

1. Identification of groups of vents (magma chambers) by utilizing geochemical data.
2. Tephra layer correlation based on both lithostratigraphic (physical) and geochemical data.

Most of the tasks related to identifying magma chambers and correlating tephra layers are currently performed manually. Nevertheless, the complexity and amount of data are ever increasing, thus it is becoming increasingly necessary to provide geologists with an aid in processing the information. This system is not supposed to replace a geologist. Geologists are deeply integrated into the processing. Geologist domain knowledge is utilized to:

- Select a relevant set of chemical elements to be considered in the process of defining magma chambers
- Constrain the number of vent groups to be consider for layer correlation
- Provide subjective opinion about qualitative stratigraphic layer attributes
- Supply a limited training set (correlated layers) for the layer recognition process

Fig. 3 shows the information flow in the intelligent system for interpreting eruption patterns. The system is built within the framework of the Transferable Belief Model (TBM) [35, 36], which is a model to represent and combine quantified beliefs based on the belief function theory developed by Shafer [34]. The TBM assumes the existence of basic belief mass $m(A) \in [01]$, where $\Theta = \{\theta_1, \ldots, \theta_n\}$ is a frame of discernment and $A \subseteq 2^{\Theta}$. $m(A)$ exists independently of any probabilistic model and represents "the amount of belief that specifically supports that the actual value of the variable on which beliefs are expressed belongs to A, and that supports nothing more specific due to a lack of information, but that might support any strict subset of A if further information justifies it" [36]. The TBM works under the open world assumption and does not assume, as do other models, that the set of hypotheses in the frame of discernment is exhaustive and that $m(\emptyset) = 0$. Several basic belief masses based on independent pieces of evidence can be combined by the so-called unnormalized Dempster rule. The TBM defines a two-level structure composed of a credal level where beliefs are estimated, and a pignistic level where decisions are made. Beliefs at the credal level are quantified by belief functions. When a decision has to be made, beliefs are transformed into pignistic probabilities using the so-called pignistic transformation. Utilization of the TBM allows for dealing with the lack of statistical data as well as uncertainty, vagueness, and imprecision inherited in the problem of rock characterization.

The following two subsections will discuss the utilization and benefits of decision fusion to the processing presented in Fig. 2.

2.2. Decision Fusion for Tephra Layer Correlation

As mentioned in the introduction, both geochemical (a percentage of each chemical element in the rocks selected by an expert) and stratigraphic characteristics, such as the fraction of pumice, grading, zoning, thickness, and size of large pumice or lithic fragments can provide information for recognition of stratigraphic

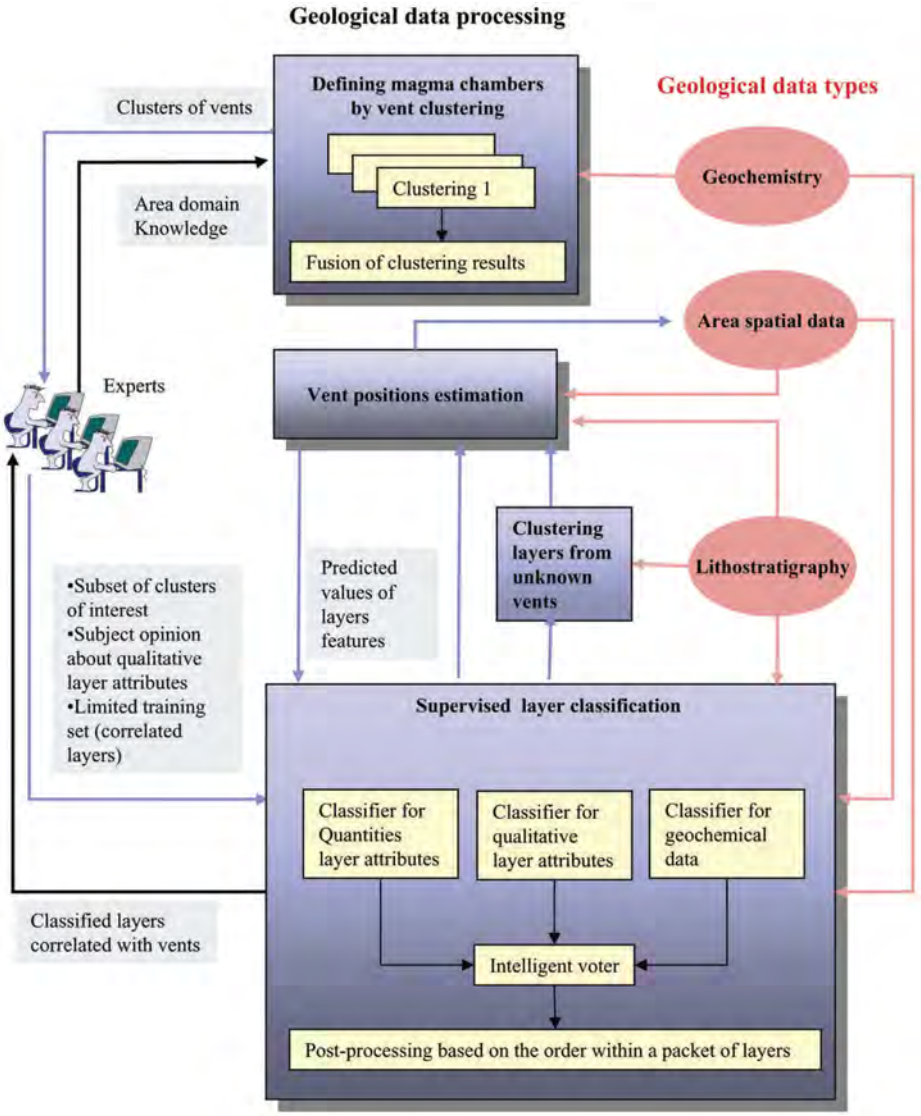


Fig. 2. Information flow in the intelligent system for interpreting eruption pattern.

eruption patterns. In many simple cases, it is straightforward for the geologist to manually categorize different layers based on these features. However, utilization of these features for tephra layer correlation faces significant challenges once the data become more complex. First, it is usually costly to collect information on past eruptions at a large number of sites as well as conduct a thorough chemical analysis of the tephra collected at these sites; hence the number of patterns used for training of the recognition system is small as compared with the dimensionality of the feature space used for recognition. Second, the features under consideration are heterogeneous and require different processing. While the chemical composition of a layer depends mostly on the magma source and characterizes a group of vents a layer could have been erupted from, the lithostratigraphic features depend on the dynamic characteristic of each eruption and geographical location relative to a single vent among many, the locations and number of which are not known at the beginning. In addition, although some of the lithostratigraphic tephra characteristics can be obtained by direct measurements, others represent vague subjective expert opinions about their values. The nonmeasurable features comprise descriptive depositional features such as zoning (e.g., more pumice on bottom than on top) and grading (e.g., bigger particles on bottom than on top) and are usually expressed in the linguistic form. For example, bedding can be characterized as reversely graded, graded, planar bedded, massive, and cross-bedded. These characteristics are subjective and uncertain since there is not always a sharp difference between, e.g., massive and graded characteristics. In many cases the uncertainty is so high that an expert may have multiple opinions about the characteristics and can assign more than one linguistic value with a certain confidence. For example, an expert could say that grading is massive with confidence 0.3 and graded with confidence 0.4. As a result it is necessary to combine the levels of confidence assigned to the value of each characteristic to convert multiple opinions about

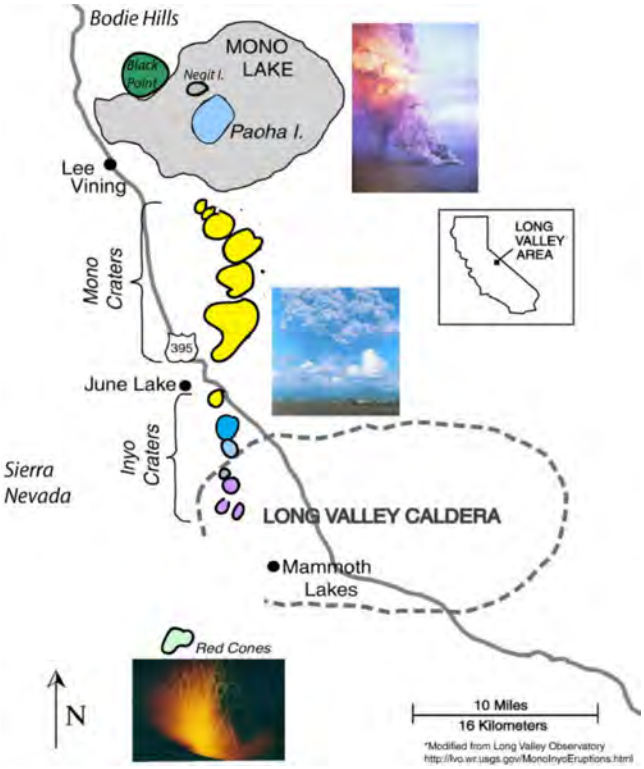


Fig. 3. Location of the Mono-Inyo Craters, showing the different types of eruptions (inset photographs) and geochemical groups of the rocks (represented by different volcano coloration).

the value of the attribute into one value to be used in a classifier [32, 46]. The measurable features comprise the data types related to eruption size in a loose sense, for example, maximum pumice, maximum lithic, and bed thickness. Their values are imprecise and depend on the expertise of the geologist doing the measurements. Finally, experts are not always sure about the labels of the patterns in the training set so the labels are uncertain.

All this calls for an iterative processing, which can deal with a small training set, heterogeneous features, uncertainty, vagueness, and imprecision characterizing features, as well as the lack of information about the number and location of the vents. To deal with insufficient number of training patterns and heterogeneous rock features a combination of the decisions of several classifiers is considered, with each of them built to treat different types of features. The classifiers are combined via an intelligent voter designed within the framework of the TBM, which is appropriate for dealing with the lack of statistical data, vague and uncertain subjective expert opinions about non-measurable tephra features as well as the open world assumption that a layer in question can be classified not only into known layers but also into a layer that has not yet been observed.

The result of the classifier combination in an intelligent voter is used for either assigning a pattern in question to a particular known class or to label this pattern as unknown. The patterns are labeled as unknown if $m(\emptyset) \geq \sum_{A \subseteq 2^\Theta} m(A)$. Otherwise the pattern is assigned to a particular known layer under consideration based on pignistic probabilities. The patterns classified either as unknowns or belonging to a certain known class can be used for augmenting the training set. The problem with patterns classified as unknown is that we cannot say whether they belong to the same class. To overcome this problem we will perform clustering of all such layers to identify the number of possible unknown classes and which unclassified layers belong to the same class.

Some preliminary results showing feasibility of utilizing the belief model and decision fusion for tephra layer correlation are presented in [8]. In [8] pairwise recognition of layers based on lithostratigraphic features was performed by combining two neural network classifiers within the framework of the TBM. Input to the first neural networks represented layer characteristics related to eruption size (maximum pumice, maximum lithic, and bed thickness). Input to the second neural network comprised descriptive deposit features (fraction of pumice and grading). Utilization of decision fusion allowed us to deal with an extremely small number of patterns available for training of the classifiers.

The results of the lithostratigraphic layer correlation can be improved by incorporating the knowledge of which batch of magma is responsible for the layer into the system. This knowledge can be obtained by processing geochemical characteristics of the rocks, as described in the next subsection.

2.3. Ensemble of Clusters for Defining Magma Chambers

Information about the size and location of magma chambers independently contributes to the knowledge about the age and eruption pattern in an area. It also provides important constraints for the process of lithostratigraphic layer correlation. Geochemistry has been used extensively to attempt to delineate separate magma batches and sources that have persisted for over a million years (see, e.g. [2, 3, 16, 18, 37]).

Information processing of geochemical data for defining location and size of magma chambers can be of two types. One of these types is based on statistical methods such as chi-square hypothesis testing utilizing a quasi-Euclidian distance between mean chemical compositions of pairs of tuff outcrops [25, 26], and factor analysis [21, 44] based on a correlation matrix for extracting common "factors" from a given data set. As summarized in [39], the use of correlation coefficients requires a multivariate normal distribution for all the input data [30]. This condition is almost never fulfilled when working with geochemical data [29]. Furthermore, geochemical data are "closed" data, which sum up to a constant. Hence multivariate statistical methods may deliver biased results [1, 40]. While these methods may provide insight into the underlying structure

of a data set, their use may require further analysis to identify distinct groups [39].

A more attractive method for generating correlations among tephra and dome-rock samples is clustering, aimed at discovering structure in a given set of P-dimensional feature vectors by organizing them into k groups based on their similarity in the feature space. Clustering is an unsupervised technique, which does not require any additional information, neither on which pattern belongs to which class, nor on which subgroup of patterns belongs to the same class.

One of the problems with utilizing clustering for definition of magma chambers is the existence of many different clustering methods, which may produce different results. There are two major classes of clustering methods: relational clustering and partitioning (see, e.g. [17, 10]). Relational clustering assigns patterns to a cluster based on a similarity matrix, e.g., distances between the patterns based on their attributes or subjective expert judgments. The most popular similarity based algorithms are hierarchical algorithms [17], which are used to either merge smaller clusters into larger ones, or to split larger clusters. The result of the algorithm is a tree of clusters, called dendrogram, which shows how the clusters are related. By cutting the dendrogram at a desired level, a clustering of the data items into disjoint groups is obtained.

Partitioning directly decomposes the data set into a set of disjoint clusters. More specifically, partitioning attempts to determine K partitions that optimize a certain criterion function in an iterative procedure. Very popular partitioning algorithms are the k-means algorithm and its modification, fuzzy, possibilistic and evidential k-means [6, 17, 22], in which observations can be assigned to multiple clusters with various degrees, participation or membership coefficients. A different partitioning approach is based on probability density function estimation using Gaussian mixtures. The specification of the parameters of the mixture is based on the expectation-maximization (EM) algorithm [9].

In spite of the advantages of clustering methods as compared to multivariate statistical methods, utilization of the former for assisting in making decisions on correlation of igneous rocks and volcanic deposits presents certain problems, since clustering suggests several possible groupings, which makes it difficult to select the best one to be used for decision making. Domain knowledge about the data to be clustered and correctness of the result is very limited or does not exist. Indeed, there are many clustering algorithms available and the result of these algorithms depends on selection of algorithm parameters, such as the number of clusters to consider, the proximity measure used in the objective function, or the selection of the distance measure for building the similarity matrix and the termination point for the hierarchical algorithms. The nature of geochemical data (neither normal nor log-normal, strongly skewed, often multi-modal data distributions) makes cluster analysis results strongly dependent on the clustering algorithm selected, which allows for employing them as an "exploratory data analysis tool" [40].

In many cases a cluster validity measure may be used to select a clustering algorithm and a set of parameters producing the best clustering result. At the same time, there are several cluster validity measures reported in the literature (see, e.g. [4, 7, 27]), and selecting the best one greatly depends on the data to be clustered.

Recently, inspired by the success of combination of decisions of supervised classifiers, several papers on combining clustering results have appeared (see e.g. [11, 12, 14, 20, 38]). They demonstrate the superior performance of combinations in discovering clusters of arbitrary shape and size as compared to the performance of a single clustering algorithm. Thus the incorporation of the combination of clustering results into the process of defining magma chambers makes the clustering results more accurate and thereby leads to improved interpretation.

The next section will discuss the problem of combining the results of several clustering algorithms in more detail and introduce a new combinational method designed in the framework of the Transferable Belief Model.

3. COMBINING SEVERAL CLUSTERING RESULTS

3.1. Related Work

The general problem of combining several partitions can be formulated the following way. Let $P = \{p_1,\ldots,p_N\}$ be an ordered set of N patterns to be clustered and let C be a set of K different partitions of P: $C = \{C_1,\ldots,C_K\}$ with each partition C_k containing n_k clusters. The goal is to find a combined partition C containing n clusters.

The problem of cluster combination is more difficult than the problem of combining several supervised classifiers. First, it is not clear which cluster label in one partition corresponds to a cluster label in another partition since the cluster numbers have only symbolic meaning. Second, in the case of fusion of supervised classifiers the number of classes, to which patterns have to be assigned is usually known and is the same for all classifiers under consideration. This is not usually the case when several clustering results are combined.

Several approaches to the problem of combining clustering results are reported in the literature. One approach employs the fact that clustering is an optimization procedure based on a specific clustering criterion and regards clustering combination as a technique that constructs and processes multiple clustering criteria rather than a single criterion [45]. One of the disadvantages of this type of method is that it requires access to the original features of the data points, which are not always available. Another one is that it requires a set of a priori parameters to perform the combination.

Another clustering combination method utilizing original pattern features is described in [22]. This method involves two steps. The first step creates credal partitions via an optimization of an objective function defined within the framework of the belief functions. The second step performs a direct combination of credal partitions obtained with the Dempster rule. This method also requires the knowledge of the number of clusters in the resulting partitions.

A different type of method combines clustering results via a consensus function, which maps multiple clustering results to a final partitioning. The advantage of such methods is that they generally do not require the knowledge of the original features of the patterns to be clustered but use stored results of clustering participating in combination. The authors of [23] identified four major classes of approaches to building consensus functions: Voting or direct "re-labeling" methods, Mixture models, Hypergraph methods, and information theory-based methods.

Direct "re-labeling methods" [10, 13, 41] involve two steps. The first step is designed to solve the label correspondence problem or the problem of finding the best permutations, which is a major difficulty in obtaining a consensus-based combination. The second step combines the clustering results of correspondent clusters. In [13, 41] the problem of finding the best permutation can be rewritten as a weighted bipartite matching problem solved by the Hungarian method. The majority [13] or plurality [41] voting is exploited for finding the final partitions.

In [10] a direct re-labeling method is applied to the combination of the results of the evidential clustering method (EVCLUS). The EVCLUS is an approach to clustering proximity data within the framework of the Dempster-Shafer theory of evidence [34]. The EVCLUS results in a so-called "credal partition," which provides an assignment of objects to each possible subset of classes. Combination of credal partitions is obtained by the conjunctive rule of combination from evidence theory. The best match between partitions is found by minimizing the degree of conflict between different permutations.

Hypergraph-based methods represent clusters as hyperedges of a graph while the nodes correspond to the objects to be clustered. The problem of consensus clustering is then reduced to finding the minimum-cut of a hypergraph. An example of the hypergraph-based algorithms can be found in [38], in which three effective algorithms for solving the k-way min-cut partitioning problem are proposed. A probabilistic model of consensus, which uses a finite mixture of multinomial distributions in a space of clusterings, is given in [42]. An example of the information theoretic approaches is given in [43], in which consensus functions related to intra-class variance criteria are described.

The majority of the consensus methods need a priori knowledge about the data such as a predefined number of clusters in the combination results. The direct re-labeling methods often also require the same number of clusters in each partition used in combination. In many cases these models involve optimization that makes them computationally expensive, especially when applied to a large number of patterns.

A more attractive combination approach resembling the hard voting scheme for classifier fusion is described in [11, 12, 15]. This "pairwise approach" utilizes a square co-association matrix $\Omega = (\omega_{ij})$, $i, j = \overline{1,N}$, (N is the number of patterns), which represents an average of consensus matrices built for each partition C_k:

$$\Omega = 1/N \sum_{k=1}^{K} \Omega^k \tag{1}$$

where $\Omega^k = (\omega_{i,j}^k)$ and $\omega_{ij}^k = 1$ if p_i and p_j are in the same cluster in partition C_k and 0 otherwise. Elements ω_{ij} of the consensus matrix are used to define whether patterns p_i and p_j belong to the same cluster in the resulting partition C. Similar to a hard voting algorithm for decision fusion in supervised classification ω_{ij} is considered as evidence of the consensus between partitions under consideration. Matrix Ω can be used as input to any algorithm based on similarity patterns, for example, in the single link and the average-link methods [11].

The pairwise approach has become quite popular despite the fact that it can be rather expensive computationally if the number of patterns is very large. There are several reasons for the popularity: it is straightforward, can work with a different number of clusters in each partition to be combined, and does not require defining a priori the number of clusters in the ensemble. At the same time the pairwise approach described above has the same drawback as any hard voting algorithm, namely it does not consider the level of confidence that p_i and p_j are in the same cluster for each partitioning.

The next subsection introduces an evidential voting method representing a soft modification of the pairwise approach, which takes into account a level of certainty of assigning each pattern to a certain cluster. This method represents a refined version of the method described in [31].

3.2. Evidential Method for Combining Clustering Results

The method described in this section utilizes the stored outputs of partitioning to be combined for building an evidential consensus matrix, which represents a new natural pattern proximity measure to be used in any clustering algorithm based on the distance between patterns. We start with the description of this method for combining the results of the fuzzy k-means algorithms.

Let us consider a frame of discernment $\Theta = \{\theta_1, \theta_2\}$, where θ_1 and θ_2 are the hypotheses that each pair of patterns p_i and p_j belongs to the same or different clusters. First, we need to represent beliefs in

each hypothesis for each partition C_k. The beliefs for each partitioning have to preserve the assignment of a pattern to a cluster based on the maximum membership.

Let $U^k = (u^k_{it})$ be a membership matrix for partition C_k defining the level of participation of pattern p_i in cluster t. The values of U^k are used to produce a collection of basic probability assignments that represent evidence pro and against each hypotheses θ_i, $i = 1,2$. Let $t = \arg\max_m(u^k_{im})$ and $l = \arg\max_m(u^k_{jm})$. If $t = l$ patterns p_i and p_j are assigned to the same cluster in partitioning C_k, a degree of support for this assignment can be represented by $(1 - |u^k_{it} - u^k_{jl}|)$, which reflects our belief that the smaller is the difference between the respective values of the membership matrix, the higher is the evidence that p_i and p_j belong to the same cluster. The reliability of this assignment for each partition can be variable, and we need to use a discounted degree of support with reliability coefficients for patterns p_i and p_j respectively:

$$\begin{aligned} R^k_i &= (n^k \cdot u^k_{it} - 1)/(n^k - 1) \qquad \text{and} \\ R^k_j &= (n^k \cdot u^k_{jl} - 1)/(n^k - 1). \end{aligned} \tag{2}$$

The reliability coefficients represent the difference between the maximum coefficient defining the assignment of patterns p_i and p_j to clusters t and l respectively and an average of the rest of the membership coefficients, and reflect the level of confidence in this assignment. The discounted degree of support defines a simple support function with focus θ_1:

$$\begin{aligned} m^k_{1_{ij}}(\theta_1) &= (1 - |u^k_{it} - u^k_{jl}|) \cdot R^k_i \cdot R^k_j \\ m^k_{1_{ij}}(\Theta) &= 1 - m^k_{1_{ij}}(\theta_1). \end{aligned} \tag{3}$$

Similarly, if $t \neq l$, we can define degrees of support for assignment of patterns p_i and p_j to different clusters: $|u^k_{it} - u^k_{jt}|$ and $|u^k_{il} - u^k_{jl}|$, and corresponding discounted separable support function with focus θ_2:

$$\begin{aligned} m^k_{2_{ij}}(\theta_2) &= 1 - (1 - R^k_i|u^k_{it} - u^k_{il}|) \\ &\quad \cdot (1 - R^k_j|u^k_{jl} - u^k_{jt}|) \\ m^k_{2_{ij}}(\Theta) &= 1 - m^k_{2_{ij}}(\theta_2). \end{aligned} \tag{4}$$

For each pair of patterns p_i and p_j a set of partitions under consideration $\{C_k\}$ can be considered as a union of two subsets: $\{C_k\} = \{C^{ij}_{\text{same}}\} \cup \{C^{ij}_{\text{diff}}\}$, where $\{C^{ij}_{\text{same}}\}$ is a subset of K_{same} partitions in which p_i and p_j belong to the same clusters, while $\{C^{ij}_{\text{diff}}\}$ is a subset of K_{diff} partitions in which p_i and p_j belong to the different clusters. The combination of all support functions with focus θ_1 defined by members of subset $\{C^{ij}_{\text{same}}\}$ is

$$\begin{aligned} m^{\text{same}}_{ij}(\theta_1) &= 1 - \prod_{k:C_k \subset \{C^{ij}_{\text{same}}\}}^{K_{\text{same}}} (1 - m^k_{1_{ij}}(\theta_1)) \\ m^{\text{same}}_{ij}(\Theta) &= 1 - m^{\text{same}}_{ij}(\theta_1). \end{aligned} \tag{5}$$

The combination of all support function pro hypothesis θ_1 defined by members of subset $\{C^{ij}_{\text{diff}}\}$ is

$$\begin{aligned} m^{\text{diff}}_{ij}(\theta_2) &= 1 - \prod_{k:C_k \subset \{C^{ij}_{\text{diff}}\}}^{K_{\text{diff}}} (1 - m^k_{2_{ij}}(\theta_2)) \\ m^{\text{diff}}_{ij}(\Theta) &= 1 - m^{\text{diff}}_{ij}(\theta_2). \end{aligned} \tag{6}$$

The result of combination of $m^{\text{same}}_{ij}(\theta_1)$ and $m^{\text{diff}}_{ij}(\theta_1)$ with the normalized Dempster rule defines the combined belief that patterns p_i and p_j belong to the same or different clusters:

$$\begin{aligned} \text{bel}_{ij}(\theta_1) &= m^{\text{same}}_{ij}(\theta_1) \cdot (1 - m^{\text{diff}}_{ij}(\theta_2))/C \\ \text{bel}_{ij}(\theta_2) &= m^{\text{diff}}_{ij}(\theta_2) \cdot (1 - m^{\text{same}}_{ij}(\theta_1))/C \\ \text{bel}_{ij}(\Theta) &= (1 - m^{\text{diff}}_{ij}(\theta_2)) \cdot (1 - m^{\text{same}}_{ij}(\theta_1)/C \end{aligned} \tag{7}$$

where $C = 1 - m^{\text{diff}}_{ij}(\theta_2) \cdot m^{\text{same}}_{ij}(\theta_1)$ is a normalizing coefficient and bel(Θ) represents our ignorance. The combined similarity matrix e_{ij} is defined by the corresponding pignistic probability, and the combined partition C can be obtained by using any hierarchical algorithm based on the similarity matrix. For the crisp k-means clustering algorithm, the values of U^k used in equations (2)–(7) can be replaced by $1 - d^k_{im}$, where d^k_{im} is a relative distance between each pattern p_i and all cluster centers.

Multiple partitions can be created by considering:

1. A different subset of chemical elements characterizing rocks.
2. A different set of features characterizing patterns, for example, features obtained by different chemical analysis methods (X-ray diffraction or microprobe) or different chemical components can be considered for clustering volcanic vents.
3. Various algorithm parameters such as initial number of clusters and cluster centers. Although very popular, this method of creating multiple partitions has a serious drawback since it can generate an infinite number of results. Here we propose the consideration of different cluster validity measures instead. Each cluster validity measure allows for selecting the best set of initial parameters and reduces the number of partitions to consider.
4. Various distance measures used in the algorithm objective functions, for example Euclidian or Mahalanobis distances.

The next section describes experiments and results of application of the combination method introduced above to understand the number of separable magma batches currently active and underlying the Mono-Inyo Craters [18], a rhyolitic volcanic field with a complicated eruption history.

TABLE I
Fuzzy Clustering Results Corresponding to Optimal Values of Different Validity Indices and Results of their Ccombinations as Compared with Partitioning Suggested by Experts (average over 7 trails)

Method	Rezaee Index	Fukuyama-Sugeno Index	Xie-Beni Index	Combination with Hard Consensus Matrix	Combination with Evidential Consensus Matrix
Average % of patterns grouped incorrectly	38.74%	35.93%	36.58%	28.79%	20.56%

TABLE II
Fuzzy Clustering Results Corresponding to Optimal Values of Different Validity Indices and Results of their Combinations as Compared with Partitioning Suggested by Experts ("knowns" only)

Method	Rezaee Index	Fukuyama-Sugeno Index	Xie-Beni Index	Combination with Hard Consensus Matrix	Combination with Evidential Consensus Matrix
Average % of patterns grouped incorrectly	37.88%	33.33%	37.88%	33.33%	19.70%
# of clusters	9	3	9	9	8

4.3. Experiments and Results

The Mono-Inyo Craters volcanic chain cuts across ancient Long Valley caldera, and consists of volcanic domes, craters and lava flows that stretch for 50 km north-south, subparallel to the eastern front of the Sierra Nevada (Fig. 3). Although there may be some older events, almost all eruptions within the chain occurred less than 50,000 years ago. Because of the variety of magma and eruption types, and the migration of vents in time and space, it is nontrivial to discern patterns of eruption behavior.

We conducted experiments with patterns from the Mono-Inyo Craters characterized by the concentration of 15 chemical elements, representing an indication of originating from different magma chambers. Magma chambers are inaccessible to direct observation and the information on ground truth for determining the accuracy of the result of clustering of geochemical data is rarely available. However the ground thruth for the partitioning of some patterns can be established by checking the distribution of the resulting clusters on a map and evaluating this distribution against known properties of the survey area [40] with further confirmation of the result by consensus of a group of experts.

Our database included 66 patterns, for which partitioning into 7 clusters was confirmed by expert consensus ("knowns") and 68 patters without annotated class labels ("unknowns"). The data used for clustering comprised 91 patterns from the database, which included 66 "knowns" and 25 randomly selected "unknowns." The performace of the proposed combination method was evaluated based on an *accuracy* score [15] computed as the proportion of the correctly partitioned "known" patterns.

The experiments were conducted with combination of fuzzy k-means clustering, which is more appropriate for clustering geochemical data for defining magma chambers since the level of participation of each pattern in each cluster allows for consideration of a degree of partial membership, which can be construed to be a measure of the degree of mixing of magma. The maximum number of clusters considered in all the experiments was 9. The clustering results selected for combination corresponded to the optimum values of various cluster validity measures obtained in 25 runs each. Utilization of validity measures allowed for reducing the number of partitions to be combined and helped to avoid inclusion of very weak partitioning in the combination. The combined partition was obtained by applying the single-link method over the similarity matrix by using a natural fixed threshold of 0.5 over pignistic probability.

The results of the fuzzy k-means algorithms included in the combinations were obtained with a degree of fuzziness $m = 2$, as used in the majority of practical applications [7]. Three well known cluster validity measures were employed: Xie-Beni and Fukuyama-Sugeno indices optimizing different functions of cluster compactness and separation (see, e.g. [24], and Rezaee index based on measures of the degree of variance within each cluster [27]. The combination results were averaged over 7 trails: one with the "known" patterns only and 6 with both "knowns" and 25 "unknowns" selected randomly. The fusion results were compared with individual partitions participating in the combinations, the fusion results obtained with a similar method based on concensus matrix obtained with majority voting [11, 12], and ground truth (Tables I and II).

The results presented in Tables I and II demonstrate a superior performance of the evidential combination method as compared with the results corresponding to the best values of the validity indecies as well as the hard combination method.

6. CONCLUSIONS

The paper presents applications of fusion techniques to designing an intelligent system to support geologists

in processing complex rock characteristics for interpreting eruption patterns. The paper has also discussed the system architecture and applicability and benefits of decision fusion within the framework of the Transferable Belief Model to designing of such a system.

Special attention has been paid to the description of a new evidential method for clustering combination aimed at improved vent partitioning based on geochemical features for defining the size and position of magma chambers. This soft method utilizes the belief model for building and combining the evidential consensus matrix, which represents a new natural pattern proximity measure and can be used in any clustering algorithm based on the distance between patterns.

The research reported in the paper is a step in designing an intelligent system to support geologists in eruption forecasting for volcanic chains and fields, areas that would otherwise be difficult, perhaps impossible to characterize and understand for large amounts of complex data. For a single geologist working in a complex region with large amounts of geochemical and lithostratigraphic data, the techniques provide a route to obtaining a "second opinion" about the meaning of those data. For the Mono-Inyo Craters used in the present example, these results allow us to gain confidence about our estimates regarding the number and positions of different batches of magma in the subsurface that could potentially generate a future eruption.

REFERENCES

[1] J. Aitchison
The statistical analysis of compositional data.
New York: Wiley, 1986.

[2] R. A. Bailey
Geologic map of Long Valley Caldera, *Mono-Inyo craters volcanic chain, and vicinity, eastern California, Miscellaneous Investigations Series*.
U.S. Geological Survey, 1989, 11.

[3] R. A. Bailey, G. B. Dalrymple and M. A. Lanphere
Volcanism, structure, and geochronology of Long Valley Caldera, Mono County, California.
Journal of Geophysical Research, **81** (1976), 725–744.

[4] S. Bandyopadhyay and U. Maulik
Nonparametric genetic clustering: Comparison of validity indices.
IEEE Transactions on Systems, Man, Cybernetics—Part C: Applications and Reviews, **31**, 1 (Feb. 2001), 120–125.

[5] M. Battaglia and P. Segall
The interpretation of gravity changes and crustal deformation in active volcanic areas.
Pure and Applied Geophysics, **161**, 7 (July 2004), 0033–4553.

[6] J. C. Bezdek
Pattern Recognition with Fuzzy Objective Function Algorithms.
New York: Plenum Press, 1981.

[7] J. C. Bezdek and N. R. Pal
Some new indexes of cluster validity.
IEEE Transactions on Systems, Man, Cybernetics—Part C: Cybernetics, **28**, 3 (June 1998), 301–315.

[8] M. Bursik and G.Rogova
Use of neural networks and decision fusion for lithostratigraphic correlation with sparse data, Mono-Inyo Craters, California.
Computers & Geosciences, **32**, 10 (Dec. 2006), 1564–1572.

[9] A. P. Dempster, N. M. Laird and D. B. Rubin
Maximum likelihood from incomplete data via the EM algorithm.
Journal of the Royal Statistical Society, **B**, 39 (1977), 1–38.

[10] T. Denoeux and M. Masson
EVCLUS: Evidential clustering of proximity data.
IEEE Transactions on Systems, Man and Cybernetics, Part C, **34**, 1, (Feb. 2004), 95–109.

[11] A. L. Fred
Finding consistent clusters in data partitions.
In J. Kittler and F. Roli (Eds.), *Multiple Classifier Systems*, LNCS 2096, Springer, 2001, 309–318.

[12] A. Fred and A. K. Jain
Combining multiple clusterings using evidence accumulation.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**, 6 (June 2005), 835–850.

[13] B. Fischer and J. M. Buhmann
Bagging for path-based clustering.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **25**, 11, (Nov. 2003), 1411–1415.

[14] J. Ghosh
Multiclassifier systems: Back to the future.
In J. Kittler and F. Roli, (Eds.), *Proceedings of the Third International Workshop Multiple Classifier Systems*, LNCS, vol. 2364, Springer, 2002, 1–15.

[15] D. Greene, A. Tsymbal, N. Bolshakova and P. Cunningham
Ensemble Clustering in Medical Diagnostics.
Technical Report TCD-CS-2004-12, Dept. of Computer Science, Trinity College, Dublin, Ireland, 2004.

[16] W. Hildreth
Volcanological perspectives on Long Valley, Mammoth Mountain, and Mono Craters: Several contiguous but discrete systems.
Journal of Volcanology and Geothermal Research, **136** (2004), 169–198.

[17] A. K. Jain, M. N. Murty and P. J. Flynn
Data clustering: A review.
ACM Computing Surveys, **31**, 3, (Sept. 1999), 264–323.

[18] P. C. Kelleher and K. L. Cameron
The geochemistry of the Mono Craters-Mono Lake islands volcanic complex, eastern California.
Journal of Geophysical Research, B, Solid Earth and Planets, **95**, 17 (1990), 643–659.

[19] R. Krishnapuram, H. Frigui and O. Nasraoui
Fuzzy and possibilistic shell clustering algorithm and their application to boundary detection and surface approximation.
IEEE Transactions on Fuzzy Systems, **3** (1995) 29–60.

[20] L. I. Kuncheva and D. P. Vetrov
Evaluation of stability of k-means cluster ensembles with respect to random initialization.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **28**, 11 (Nov. 2005).

[21] R. W. Le Maitre
Numerical Petrology.
Amsterdam: Elsevier, 1982.

[22] M-H Masson and T. Denoeux
ECM: An evidential version of the fuzzy c-means algorithm.
Pattern Recognition, **41** (2008), 1384–1397.

[23] B. Minaei-Bidgoli, A. Topchy and W. F. Punch
A comparison of resampling methods for clustering ensemble.
International Conference on Machine Learning; Models, Technologies and Applications (MLMTA04), June 2004.

[24] N. R. Pal and J. C. Bezdek
On cluster validity for the fuzzy c-means model.
IEEE Transactions on Fuzzy Systems, **3** (Apr. 1995), 370–376.

[25] M. E. Perkins, W. P. Nash, F. H. Brown and R. J. Fleck
Fallout tuffs of Trapper Creek, Idaho—A record of Miocene explosive volcanism in the Snake River Plain volcanic province.
GSA Bulletin, **107**, 12 (1995), 1484–1506.

[26] M. E. Perkins and W. P. Nash
Explosive silicic volcanism of the yellow stone hotspot: The ash fall tuff record.
GSA Bulletin, **114**, 3 (2002), 367–381.

[27] M. R. Rezaee, B. P. F. Lelieveldt and J. H. C. Reiber
A new cluster validity index for the fuzzy c-mean.
Pattern Recognition Letters, **19** (1998), 237–246.

[28] M. F. Reed, R. C. Bartholomay and S. S. Hughes
Geochemistry and stratigraphic correlation of basalt lavas beneath the Idaho Chemical Processing Plant, Idaho National Engineering Laboratory.
Environmental Geology, **30**, 1/2 (1997), 108–118.

[29] C. Reimann and P. Filzmoser
Normal and lognormal data distribution in geochemistry: Death of a myth. Consequences for the statistical treatment of geochemical and environmental data.
Environmental Geology, **39**, 9 (2000), 1001–1014.

[30] C. Reimann, P. Filzmoser and R. G. Garrett
Factor analysis applied to regional geochemical data: Problems and possibilities.
Applied Geochemistry, **17** (2002), 185–206.

[31] G. Rogova, M. Bursik and S. Hanson-Hedgecock
Interpreting the pattern of volcano eruptions: Intelligent system for tephra layer correlation.
In *Proceedings of the FUSION'2007—10th Conference on Multisource Information Fusion*, 2007.

[32] G. Rogova and P. Losiewicz
Hybrid approach to multi-attribute decision making.
In *Proceedings of the FUSION'99 2nd Conference on Multisource-Multisensor Information Fusion*, 1999, 902–909.

[33] G. Rogova, P. Scott and C. Lollett
Higher level fusion for post-disaster casualty mitigation operations.
In *Proceedings of the 8th International Conference on Multisource Information Fusion* (FUSION'2005), 2005.

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

[35] P. Smets and R. Kennes
The transferable belief model.
Artificial Intelligence, **66** (1994), 191–243.

[36] P. Smets and B. Ristic
Kalman filter and joint tracking and classification based on belief functions in the TBM framework.
Information Fusion, **8**, 1 (Jan. 2007), 16–27.

[37] R. S. J. Sparks, M. Bursik, S. Carey, J. S. Gilbert, L. S. Glaze, H. Sigurdsson and A. W. Woods
Volcanic Plumes.
London: Wiley, 1997.

[38] A. Strehl and J. Ghosh
Cluster ensembles—A knowledge reuse framework for combining multiple partitions.
Journal of Machine Learning Research, **3** (2002), 583–618.

[39] S. K. Swanson, J. M. Bahr, M. T. Schwar and K. W. Potter
Two-way cluster analysis of geochemical data to constrain spring source water.
Chemical Geology **179**, 1–4 (2001), 73–91.

[40] M. Templ, P. Filzmoser and C. Reimann
Cluster analysis applied to regional geochemical data: Problems and possibilities.
2006, Webpage: http://www.statistik.tuwien.ac.at/forschung/CS/CS-2006-5complete.pdf.

[41] A. P. Topchy, M. Law, A. K. Jain and A. Fred
Analysis of consensus partition in cluster ensemble.
In *Proceedings of the Fourth IEEE International Conference on Data Mining* (ICDM'04), 2004, 225–232.

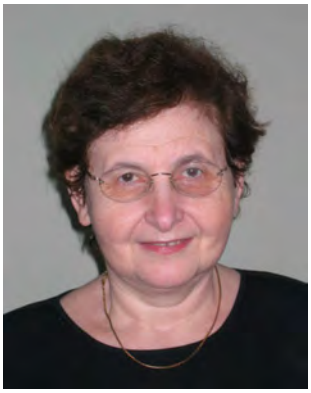
[42] Topchy, A. K. Jain and W. Punch
A mixture model of clustering ensembles.
In *Proceedings of the Fourth SIAM International Conference on Data Mining*, 2003, 379–390.

[43] A. Topchy, A. K. Jain and W. Punch
Combining multiple weak clusterings.
In *Proceedings of the IEEE International Conference on Data Mining* (ICDM), 2003.

[44] Lin Xiaotong, Li Weiran and Du Shujie
Stratigraphic division based on minor elements in the sediment core Q43 from the outer shelf of the East China Sea: Approached through correspondence analysis and fuzzy cluster analysis.
Journal of Ocean University of China (English Ed.), **4**, 2 (2005), 131–136.

[45] Y. Qian and C. Y. Suen
Clustering combination method.
In *Proceedings of the 15th International Conference on Pattern Recognition*, 2 (Sept. 2000), 732–735.

[46] J.-B. Yang and N. G. Singh
An evidential reasoning approach for multiple-attribute decision making with uncertainty.
IEEE Transactions on Systems, Man, and Cybernetics, **24**, 1 (1994), 1–78.



Galina Rogova received her M.S. and Ph.D. in Moscow, Russia.

Currently she is a consultant on behalf of Encompass Consulting, which she founded in 2000. Galina Rogova is also an adjunct professor at the State University of New York at Buffalo. Her research interests include information fusion, machine learning, pattern recognition, reasoning and decision making under uncertainty, and image analysis and understanding. Dr. Rogova has worked on a wide range of problems such as interpreting the pattern of volcano eruption, modeling decision-maker preferences under uncertainty, designing a reinforcement learning system for the distributed environment, target recognition, situation and threat assessment, computer-aided radiology, Optical Character Recognition, and Intelligent Transportation Systems. She published numerous papers and edited two books. Her research has been funded by the AFRL, the NSF, the NIH, the NGA, the US and NYS Departments of Transportation, and the DRDC, Canada-Valcartier. She is a member of the International Society of Information Fusion and the American Association for the Advancement of Science.

Marcus Bursik he received the Ph.D. degree from Caltech in 1989. He studied geology and geophysics.

Since 2000 he has been full professor of geology in the Department of Geological Sciences, University at Buffalo, State University of New York. He has held visiting professor and scientist positions at Universidad Nacional Autonomo de Mexico, University of Bristol, University of Cambridge and NASA's Jet Propulsion Laboratory. He has published more than 80 papers and two books in the field of volcanology. His main interest is volcanic plume dynamics and tephrochronology, especially in the Mono-Inyo Craters, California. He is a member of the American Geophysical Union and a life member of the International Association of Volcanology and Chemistry of the Earth's Interior.

Sara Hanson-Hedgecock graduated from Rice University with a B.S. in geophysics, in 2003 and will finish her M.S. in geology at the State University of New York at Buffalo in June 2009.

Her research interests include using geophysical methods to explore the physical and mechanical properties of earthquakes and volcanoes, as well as developing numerical models of these phenomena using geological and geophysical data.

Information for Authors

The Journal of Advances in Information Fusion (JAIF) is a semi-annual, peer-reviewed archival journal that publishes papers concerned with the various aspects of information fusion. The boundaries of acceptable subject matter have been intentionally left flexible so that the journal can follow the research activities to better meet the needs of the research community and members of the International Society of Information Fusion (ISIF).

Restrictions on Publication: The International Society of Information Fusion (ISIF) publishes in JAIF only original material that has not been either published or submitted for publication in an archival journal. However, prior publication of an abbreviated and/or preliminary form of the material in conference proceedings or digests shall not preclude publication in JAIF when notice is made at the time of submission and the previous publications are properly referenced in the manuscript. Notice of prior publication of an abbreviated version of the manuscript must be given at the time of submission to the transactions. Submission of the manuscript elsewhere is precluded unless it is refused for publication or withdrawn by the author. Concurrent submission to other publications and this journal is viewed as a serious breach of ethics and, if detected, the manuscript under consideration will be rejection.

Sanctions for Plagiarism: If a manuscript is substantially similar to any previous journal submission by any of the authors, its history and some explanation must be provided. If a manuscript was previously reviewed and rejected by a different journal, the authors shall provide to the area editor copies of all correspondence involving the earlier submission. In addition, the authors must discuss the reasons for resubmission and be prepared to deliver further material if more is requested. Submission of a previously rejected manuscript is acceptable as long as the motive for submission to JAIF is explained and the authors have addressed the feedback of the earlier reviewers. The authors of a manuscript that is found to have been plagiarized from others or contain a crossover of more than 25% with another journal manuscript by any of the authors will incur sanctions by the ISIF. The following sanction will be applied for any of the above-noted infractions: (1) immediate rejection of the manuscript in question; (2) immediate withdrawal of all other submitted manuscripts by any of the authors to any of the ISIF publications (journals, conferences, workshops); (3) prohibition against all of the authors for any new submissions to ISIF publications, either individually, in combination with the authors of the plagiarizing manuscript as well as in combination with new coauthors. The prohibition shall continue for at least two years from notice of suspension.

Types of Contributions: Contributions may be in the form of regular papers or correspondence. The distinction between regular papers and correspondence is not one of quality, but of nature. Regular papers are to be a well-rounded treatment of a problem area, while a correspondence item makes one or two points concisely. In regular papers, the title, abstract, and introduction should be sufficiently informative to illuminate the essence of the manuscript to the broadest possible audience and to place the contributions in context with related work. The body of the manuscript should be understandable without undue effort by its intended audience. Correspondence should be less discursive but equally lucid. Authors should be aware that a well-written correspondence are usually published more rapidly than a regular paper.

Submission of Manuscript for Review: Manuscripts are submitted electronically via the internet at http://jaif.msubmit.net. For peer review, manuscripts should be formatted in a single column and double spaced.

General Information for JAIF Authors and Submission of Final Manuscript: General information for JAIF authors is available at http://www.isif.org/jaif.htm, and specific document preparation information for final manuscript for publication is given below and can be found at http://www.isif.org/JAIF Manuscript Preparation1.doc.

Copyright: It is the policy of the ISIF to own the copyright to the technical contributions it publishes; authors are required to sign an ISIF copyright transfer form before publication. This form appears in JAIF from time to time and is available on the web site listed above. If a copyright form is not submitted with the manuscript, it will be requested upon contribution acceptance. Publication will not take place without a completed copyright form.

Page Charges: Since ISIF has elected to support the publication of JAIF with the revenues of conferences and membership dues, page charges are not mandated for publication in JAIF. Page charges for journals of professional societies are traditional, and at some time in the future, page charges may be introduced to cover the expenses associated with the publication of JAIF. Payment of page charges will not be a prerequisite for publication.

Discloseability: The ISIF must, of necessity, assume that material submitted for publication is properly available for general dissemination to the audiences for which JAIF serves. It is the responsibility of the author, not ISIF, to determine whether disclosure of materials requires prior consent of other parties, and, if so, obtain it.

Preparation of Manuscript for Publication: After a manuscript is recommended for publication by the editorial staff, the corresponding author prepares the manuscript according to the guidelines below and submits the final version of the manuscript as a pdf file and the supporting manuscript files. Before prepar-

ing their manuscript for publication, authors should check for the more update version of the guidelines at http://jaif.msubmit.net.

All manuscripts must be submitted electronically in pdf format with the supporting files in one of the following formats: Plain TeX, AMS-TeX, LaTeX, or MS-Word with the MathType extension for equations and in-text mathematical material. The pdf file must be created with graphics resolution set to 300 dpi or greater. The author must approve the pdf file because it will be used as the basis for rectifying any inconsistencies in the supporting files. The supporting files must be prepared according to the following guidelines.

- Text files should include title, authors, affiliations, addresses, a footnote giving the dates for submission and revision and name of the editor that handed the review, and an optional footnote acknowledging a sponsor for the research. The information must be included in the text files in this order.
- Abstract must be included. The abstract must include the key words for the manuscript and include no more than 300 words for a regular paper and 150 words for a correspondence.
- Authors should number main section headings as 1, 2, 3, etc., and subsections as 1.1, 1.2, or 2.1. All headings should be typed with title format (i.e., first letter caps and lower case)—not in ALL CAPS. The typesetter will convert to all caps in final formatting as required.
- Authors should format references very carefully according to the examples given below. Style for authors' names are initials followed by last name and it must be followed precisely. References must be listed alphabetically by (i) last name of first author, (ii) initials of first author, (iii) last name of second author, (iv) initials of second author, etc. For manuscripts with common authors and order of authors, the date of publication should be used to select order with the earlier publications being list first. The names of publications must be spelled completely. In the references, the author should use only approved abbreviations that include: No., Vol., p., pp., AIAA, IEEE, IEE, and SPIE.
- Authors who use one of the TeX variants must provide a list of their macros and the files for the macros. Authors who use LaTeX must include the bbl and aux files.
- All figures must be submitted electronically as HIGH resolution (300 dpi or better) in Color graphics or GRAYSCALE (where shading is involved) or BLACK AND WHITE (if simple line art) graphics files (tif, eps, jpg, bmp). Each figure must be supplied as a separate graphics file. Graphics (or captions) should NOT be embedded in the text files. The figures must be included in the pdf file of the full article and the pdf file must be created with graphics resolution set to 300 dpi or greater.
- A separate file including all figure captions must be included.
- Each table must be submitted in a separate file. Tables (or captions) should NOT be included in text files and should be in form of DATA—rather than graphics—files.
- A separate file including all table captions must be included.
- A separate text of the biography of each author must be submitted. The text file should be less than 500 words.
- Separate graphics files of each author's photo should be provided as a grayscale graphics file or a color graphics file.

Examples of the references are alphabetized correctly and listed below.

BOOK:

[1] R. E. Blahut
Theory and Practice of Error Control Codes. Reading, MA: Addison-Wesley, 1983.

PROCEEDINGS ARTICLE:

[2] T. Fichna, M. Gartner, F. Gliem, and F. Rombeck
Fault-tolerance of spacebome semiconductor mass memories.
In Twenty-Eighth Annual International Symposium on Fault-Tolerant Computing, Digest of Papers, 1998, 408–413.

BOOK:

[3] P. K. Lala
Fault Tolerant and Fault Testable Hardware Design. Englewood Cliffs, NJ: Prentice-Hall, 1985.

WEB SITE:

[4] National Semiconductors Inc.
Homepage: http://www.national.com.

PROCEEDINGS ARTICLE:

[5] C. Paar and M. Rosner
Comparison of arithmetic architectures for reed-solomon decoders in reconfigurable hardware.
In *Proceedings of the Symposium on Field-Programmable Custom Computing Machines*, Apr. 1997, 219–225.

JOURNAL ARTICLE:

[6] N. R. Saxena and E. J. McCluskey
Parallel signature analysis design with bounds on aliasing.
IEEE Transactions on Computers, **46**, 4 (Apr. 1997), 425–438.

[7] C. I. Underwood and M. K. Oldfield
Observations on the reliability of cots-device-based solid state data recorders operating in low-Earth orbit.
IEEE Transactions on Nuclear Science, **47**, 4 (June 2000), 647–653.

WEB SITE:

[8] Xilinx Inc.
Homepage: http://www.xilinx.com.

INTERNATIONAL SOCIETY OF INFORMATION FUSION

ISIF Website: http://www.isif.org

BOARD OF DIRECTORS*

2006-2008	2007-2009	2008-2010
David Hall	Stefano Coraluppi	Steve Andler
Ivan Kadar	Jean Dezert	Subhash Challa
Elisa Shahbazian	Wolfgang Koch	Mahendra Mallick

*Board of Directors are elected by the members of ISIF for a three year term.

PAST PRESIDENTS

Darko Musicki, 2008
Erik Blasch, 2007
Pierre Valin, 2006
W. Dale Blair, 2005

Chee Chong, 2004
Xiao-Rong Li, 2003
Yaakov Bar-Shalom, 2002

Parmod Varshney, 2001
Yaakov Bar-Shalom, 2000
Jim Llinas, 1999
Jim Llinas, 1998

SOCIETY VISION

The International Society of Information Fusion (ISIF) seeks to be the premier professional society and global information resource for multidisciplinary approaches for theoretical and applied information fusion technologies.

SOCIETY MISSION

Advocate

To advance the profession of fusion technologies, propose approaches for solving real-world problems, recognize emerging technologies, and foster the transfer of information.

Serve

To serve its members and engineering, business, and scientific communities by providing high-quality information, educational products, and services.

Communicate

To create international communication forums and hold international conferences in countries that provide for interaction of members of fusion communities with each other, with those in other disciplines, and with those in industry and academia.

Educate

To promote undergraduate and graduate education related to information fusion technologies at universities around the world. Sponsor educational courses and tutorials at conferences.

Integrate

Integrate ideas from various approaches for information fusion, and look for common threads and themes—look for overall principles, rather than a multitude of point solutions. Serve as the central focus for coordinating the activities of world-wide information fusion related societies or organizations. Serve as a professional liaison to industry, academia, and government.

Disseminate

To propagate the ideas for integrated approaches to information fusion so that others can build on them in both industry and academia.

Call for Papers

The Journal of Advances in Information Fusion (JAIF) seeks original contributions in the technical areas of research related to information fusion. Authors of papers in one of the technical areas listed on the inside cover of JAIF are encouraged to submit their papers for peer review at http://jaif.msubmit.net.

Call for Reviewers

The success of JAIF and its value to the research community is strongly dependant on the quality of its peer review process. Researchers in the technical areas related to information fusion are encouraged to register as a reviewer for JAIF at http://jaif.msubmit.net. Potential reviewers should notify via email the appropriate editors of their offer to serve as a reviewer.

Additional Copies

If you need additional copies, please contact Tammy Blair at 404-407-7726 or tammy.blair@gtri.gatech.edu for pricing information.

Journal of Advances in Information Fusion Volume 3 Number 2 December 2008 ISSN 1557-6418