On Combining Classifier Mass Functions for Text Categorization

David A. Bell, J.W. Guan, and Yaxin Bi

Abstract—Experience shows that different text classification methods can give different results. We look here at a way of combining the results of two or more different classification methods using an evidential approach. The specific methods we have been experimenting with in our group include the Support Vector Machine, kNN (nearest neighbors), kNN model-based approach (kNNM), and Rocchio methods, but the analysis and methods apply to any methods. We review these learning methods briefly, and then we describe our method for combining the classifiers. In a previous study, we suggested that the combination could be done using evidential operations [1] and that using only two focal points in the mass functions (see below) gives good results. However, there are conditions under which we should choose to use more focal points. We assess some aspects of this choice from an evidential reasoning perspective and suggest a refinement of the approach.

Index Terms—Data mining systems and tools, modeling of structured, textual and multimedia data, uncertainty reasoning.

1 Introduction

Text Categorization (TC) is a technique often used as a basis for applications in document processing and visualization, Web mining, technology watch, patent analysis, etc. Assessment of different methods by experiment is the basis for choosing a classifier as a solution to a particular problem instance. No single classifier is always best [2], so, for practical purposes, we need to develop an effective methodology for combining them.

The fusion of results from multiple classifiers (for various purposes such as image classification as well as text) may generate a better classification than the individuals concerned. Different rules have been tried for this fusion, such as the product, average, and some more esoteric rules, such as Dempster's Rule [3] or information theoretic criteria [17]. Another approach is to employ a second-level classifier which uses Decision Templates [18] to combine the results—e.g., by comparing them to a template characterizing each class. This has the advantage of using *all* of the results to arrive at the final support for each class. In [18], it was shown to be superior to other methods of combination such as majority voting or naive Bayes on LandSat data. It gave similar results to Dempster's Rule for this image classification.

There are clear benefits of combining multiple classifiers based on different classification methods for TC and these have been discussed in [3], [4], [5]. Our own approach is to use a combination method for text classifiers based on Dempster's Rule for combination of evidence, as presented in [1]. We detailed an experimental study on the method in [6]. We developed [7], [8] a novel evidence structure for representing outputs from different classifiers based on the confidence values for labels, using a 2-points focused mass function (see below), which has been employed in [1], [6]. This constitutes a piece of evidence and serves the purpose of distinguishing important elements from trivial ones. We now assess some aspects of this from an evidential reasoning perspective and suggest a refinement of the approach. We generalize the approach to cover mass functions with more foci and show how to find conditions which determine when two focal points are better than 3, 3 better than 4, etc.

2 Learning Algorithms for Text Categorization

In a prototype for an EU Framework 5 project called ICONS, we implemented some existing methods for TC and added a new one called kNNModel (see below). The existing methods included the Rocchio method, the Support Vector method, and the standard kNN method.

The Rocchio method was originally developed for query expansion by means of relevance judgments in information retrieval. It has been applied to text categorization by Ittner et al. [9]. There are several versions of the algorithm and we implemented Ittner's method.

kNN is an instance-based classification method which has been effectively applied to text categorization in the past decade. In particular, it is one of the top-performing methods on the benchmark Reuters corpus [10]. Unlike most supervised learning algorithms that have an explicit training phase before dealing with any test document, kNN makes use of the *local contexts* derived from training documents to make the classification decision on a particular document.

SVM (Support Vector Machine) is a high-performance learning algorithm which has been applied to text categorization by Joachims [11]. We have integrated a version of the SVM algorithm implemented by Chang and Lin [12] in

- *D.A. Bell is with the School of Computer Science, The Queen's University of Belfast, BT7 1NN, UK. E-mail: da.bell@qub.ac.uk.*
- *J.W. Guan is with the School of Computer Science, The Queen's University of Belfast, BT7 1NN, UK, and the College of Computer Science and Technology, Jilin University, Changchun, 130012, P.R. China. E-mail: j.guan@qub.ac.uk.*
- *Y. Bi is with the School of Computer Science, The Queen's University of Belfast, BT7 1NN, UK, and the School of Computing and Mathematics, University of Ulster at Jordanstown, Co. Antrim, BT37 0QB, UK. E-mail: yaxin.bi@bt.com and y.bi@ulster.ac.uk.*

Manuscript received 19 Nov. 2004; revised 29 Mar. 2005; accepted 1 Apr. 2005; published online 18 Aug. 2005.
For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number TKDESI-0427-1104.

1041-4347/05/$20.00 © 2005 IEEE Published by the IEEE Computer Society

our prototype system for text categorization. There are two advantages of this algorithm: The first is that it has the ability to cope with the multiclass classification problem and the second is that the classified results can be expressed as posterior probabilities that are directly comparable between categories.

The kNNModel is an integration of the conventional kNN and Rocchio algorithms [13]. It improves the kNN method by not being too dependent on our choice of k. Local models are treated as local centroids for the respective categories to overcome the deficiency of misclustering some data points when linearly clustering the space of data points.

2.1 Output of Classification Methods

We now describe the classification process in an abstract manner. Let $D = \{d_1, d_2, \ldots, d_{|D|}\}$ be a training set of documents, where d is represented by a $|V|$-dimensional weighted vector and V is a set of keywords. Let $C = \{c_1, c_2, \ldots, c_i, \ldots, c_{|C|}\}$ be a set of categories, then the task of assigning predefined categories onto documents can be regarded as mapping which maps a Boolean value true (T) or false (F) to each pair $< d, c > \in D \times C$. If value T is assigned to $< d, c >$, it means that a decision is made to include document d under the category c, whereas value F indicates that document d is not under the category c.

The task of learning for text categorization is to construct such an approximation to a unknown function φ such that makes $\varphi : D \times C \rightarrow \{T, F\}$, where φ is called a *classifier*. However, given a test document d, such a mapping cannot guarantee that an assignment of the categories to the document is either true or false; instead, it is a $|C|$-dimensional vector of numeric values, denoted by $S = (s_1, s_2, \ldots, s_i, \ldots, s_{|C|})$, where $s_i = w(c_i)$ represents the relevance of the document to the list of categories in the form of similarity scores or probabilities, i.e., $\varphi(d) = \{s_1, s_2, \ldots, s_i, \ldots, s_{|C|}\}$, where the greater the score of the category, the greater the possibility of the document being under the corresponding category. It is necessary to develop a decision rule to determine a final category of the document on the basis of these scores or probabilities.

3 Handling Uncertainty

When classifying a particular document, information and knowledge (e.g., rules) pertinent to it often originate from different evidence sources and are often pervaded with uncertainty. The question arises: Is there any way we could formalize the reasoning processes or otherwise make more visible for practical application how evidence (uncertain knowledge and information) pertinent to a situation is obtained from multiple sources and combined?

We adopt an evidential approach for this. Exploitation of the different pieces of evidence usually requires combination operations such as Dempster's Rule or the *orthogonal sum* [14] to solve the Data/Information/Knowledge fusion problem.

Decision making involves finding the best supported option based on all the available evidence. One traditional approach to evidential reasoning, based on numerical methods of representing evidential supports, is the Dempster-Shafer (D-S) theory of evidence, which can make use of quantitative information available (from classifiers, in the present context).

3.1 Evidence Theory

The D-S theory of evidence has been recognized as an effective method for coping with such uncertainty or imprecision embedded in evidence used in the reasoning process. It is suited to a range of decision-making activities. The D-S theory is often viewed as a generalization of Bayesian probability theory by providing a coherent representation for *ignorance* (lack of evidence) and also by discarding the *insufficient reasoning principle*. It formulates a reasoning process as pieces of evidence and hypotheses and subjects these to a strict formal process to infer conclusions from the given uncertain evidence, avoiding human subjective intervention to some extent.

In the D-S theory, which we also refer to as evidence theory, evidence is described in terms of evidential functions. Several functions commonly used in the theory are *mass functions*, *belief functions*, *commonality functions*, *doubt functions*, and *plausibility functions*. Any one of these conveys the same information as any of the others.

Definition 1. *Let Θ be a finite nonempty set and call it the* frame of discernment. *Let $[0,1]$ denote the interval of real numbers from zero to one, inclusive: $[0,1] = \{x | 0 \leq x \leq 1\}$. A function $m : 2^\Theta \rightarrow [0,1]$ is called a* mass function *if it satisfies: $m(\emptyset) = 0$, $\sum_{X \subseteq \Theta} m(X) = 1$.*

A mass function is a *basic probability assignment bpa* to all subsets X of Θ. A subset A of a frame Θ is called a *focal element* of a mass function m over Θ if $m(A) > 0$. Note that a focal element is a subset rather than an element of Θ. The union C of all the focal elements of a mass function is called its *core*: $C = \cup_{X, m(X)>0} X$.

A function $bel : 2^\Theta \rightarrow [0,1]$ is called a *belief function* if it satisfies: $bel(\emptyset) = 0$; $bel(\Theta) = 1$; for any collection $A_1, A_2, \ldots, A_n$ $(n \geq 1)$ of subsets of Θ,

$$bel(A_1 \cup A_2 \cup \ldots \cup A_n) \geq \sum_{I \subseteq \{1,2,\ldots,n\}, I \neq \emptyset} (-1)^{|I|+1} bel(\cap_{i \in I} A_i).$$

This expression can be contrasted with conventional probability, where the inequality is replaced by an equality.

The fundamental operation of evidential reasoning, namely, the orthogonal sum of evidential functions, is known as Dempster's rule for combining evidence. Let m_1 and m_2 be mass functions on the same frame Θ. Denote $N = \sum_{X \cap Y \neq \emptyset} m_1(X) m_2(Y)$. Suppose $N > 0$, i.e., $\sum_{X \cap Y = \emptyset} m_1(X) m_2(Y) < 1$. Then, the function $m : 2^\Theta \rightarrow [0,1]$ defined by: $m(\emptyset) = 0$ and

$$m(A) = (1/N) \sum_{X \cap Y = A} m_1(X) m_2(Y)$$

for all subsets $A \neq \emptyset$ of Θ is a mass function. The mass function m is called the orthogonal sum of m_1 and m_2 and is denoted by $m_1 \oplus m_2$, and $K = 1/N$ is called the *normalization constant* of the orthogonal sum of m_1 and m_2.

Testing docs	|	classifier$_1$	+	classifier$_2$	$\rightarrow$	combined classifier
d_1	|	triplet$_{11}$	$\oplus$	triplet$_{21}$	=	result$_1$
$\vdots$	|	$\vdots$	$\oplus$	$\vdots$	=	$\vdots$
d_n	|	triplet$_{1n}$	$\oplus$	triplet$_{2n}$	=	result$_n$

Fig. 1. The procedure of combining classifiers.

3.2 A Categorization-Specific Mass Function

We now consider the problem of estimating the degrees of belief for the evidence obtained from text classifiers and the specific definitions of mass and belief functions for this domain. We then look at how to fuse multiple pieces of evidence in order to reach a final decision.

Definition 2. *Let C be a frame of discernment, where each category $c_i \in C$ is a proposition of the form: "document d is of category c_i" and let $\varphi(d)$ be a piece of evidence that indicates the strength of our confidence that the document comes from each category $c_1, c_2, \ldots, c_i, \ldots, c_{|C|}$. Then, the following mass function m is a* basic probability assignment (bpa) *to c_i for*

$$1 \leq i \leq |C| : m(\{c_i\}) = \frac{s_i}{\sum_{k=1}^{|C|} s_k} = \frac{w(c_i)}{\sum_{k=1}^{|C|} w(c_k)}.$$

This mass function bpa *expresses the degrees of beliefs in respective propositions corresponding to each category to which a given document could belong.*

We can rewrite the expression for the output information $\varphi(d)$ as $\varphi(d) = (m(c_1), m(c_2), \ldots, m(c_i), \ldots, m(c_{|C|}))$. Two or more outputs derived from different classifiers as pieces of evidence can then be combined using the orthogonal sum. In order to improve the efficiency of computing orthogonal sum operations and the accuracy of a final decision on the basis of the combined results, we have developed a new structure using a 2-points focused mass function (see below), which partitions $\varphi(d)$ into three subsets [7], [8]. This is an example of a *truncated foci mass function* and the concept can be generalized to deal with more than three subsets. Empirical evaluations have shown that it is effective and that using only the best classifiers gives good results. Some theoretical work for its validity and ability to be combined can be found in [8].

Definition 3. *Let C be a frame of discernment and*

$$\varphi(d) = (m(c_1), m(c_2), \ldots, m(c_i), \ldots, m(c_{|C|})),$$

where $|C| \geq 2$. Consider an expression of the form $Y = (A_1, A_2, A_3)$, where $A_1, A_2 \subseteq C$ are singleton, and A_3 is the whole set C. These elements are given by the formulae below. Arrange $\varphi(d)$ so that $m(c_{i_1}) \geq m(c_{i_2}) \geq \ldots \geq m(c_{i_{|C|}})$. Then, $A_1 = \{c_{i_1}\}, A_2 = \{c_{i_2}\}, A_3 = C$.

The associated 2-points focused mass function is given as follows: $m(A_1) = m(\{c_{i_1}\})$, $m(A_2) = m(\{c_{i_2}\})$, $m(A_3) = 1 - m(A_1 - m(A_2))$, i.e.,

$$m(C) = 1 - m(\{c_{i_1}\}) - m(\{c_{i_2}\}).$$

```
Result 1 (SVM):
<DOCUMENT name="37928" category =" c2">
    <CATEGORY bpa = "0.724">{c1}</CATEGORY>
    <CATEGORY bpa = "0.184">{c2}</CATEGORY>
    <CATEGORY bpa = "0.092">{c1, c2, c3, c4, c5, c6}</CATEGORY>
</DOCUMENT>
Result 1 (kNN):
<DOCUMENT name="37928" category =" c2">
    <CATEGORY bpa = "0.688">{c2}</CATEGORY>
    <CATEGORY bpa = "0.208">{c4}</CATEGORY>
    <CATEGORY bpa = "0.104">{c1, c2, c3, c4, c5, c6}</CATEGORY>
</DOCUMENT>
```

Fig. 2. Outputs produced by kNN and SVM.

We refer to the set $\{A_1, A_2, A_3\}$ as a *triplet*. Previously, we made the assumption that the categories to be assigned to a given document include only the top choice, the top second choice, or the whole of the frame in descending order for each classifier. It is then possible that the second top choice for a classifier will be ranked as the top choice when we combine multiple classifiers. This provides the rationale behind dividing $\varphi(d)$ into a triplet. We show how this can be done in Section 5 and compare it with structures having more than two foci.

4 The 2-Points Focused Combination Method

Assume that we have a set of training data and a set of algorithms, each of which can generate one or more classifiers based on the training data set chosen. We can combine outputs of different classifiers on the same testing documents using Dempster's rule of combination to make the final classification decision. Fig. 1 illustrates the process of combining the outputs of two classifiers derived from two different learning algorithms.

Consider an example where we are given two triplets (A_1, A_2, C) and (B_1, B_2, C), where $A_1, A_2, B_1, B_2 \subseteq C$, and two associated 2-points focused mass functions m_1, m_2. These are two pieces of evidence. Suppose that they are obtained from two classifiers kNN and SVM, respectively, represented in XML [7] in Fig. 2. In this example, $C = \{c_1, c_2, c_3, c_4, c_5, c_6\}$ is a frame of discernment, where c_i $(i = 1, 2, \ldots, 6)$ are document categories. That is, c_1: comp. windows.x; c_2: comp.graphics; c_3: comp.sys.ibm.pc.hardware; c_4:comp.sys.mac.hardware; c_5: comp.os.ms-windows. misc; c_6: alt.atheism.

We use triplets (A_1, A_2, C) and (B_1, B_2, C) to represent the two results, i.e.,

$$(A_1, A_2, C) = (\{c_1\}, \{c_2\}, \{c_1, c_2, c_3, c_4, c_5, c_6\});$$
$$(B_1, B_2, C) = (\{c_2\}, \{c_4\}, \{c_1, c_2, c_3, c_4, c_5, c_6\}).$$

The values of the corresponding mass functions on these propositions for document 37928 are shown in Fig. 2.

TABLE 1
Intersection Table of Two Quartets

$m_1 \oplus m_2$	$\{x\}_{m_2(\{x\})}$	$\{y\}_{m_2(\{y\})}$	$\{z\}_{m_2(\{z\})}$	$\Theta_{m_2(\Theta)}$
$\{x\}_{m_1(\{x\})}$	$\{x\}_{m_1(\{x\})m_2(\{x\})}$	$\{\}_{m_1(\{x\})m_2(\{y\})}$	$\{\}_{m_1(\{x\})m_2(\{z\})}$	$\{x\}_{m_1(\{x\})m_2(\Theta)}$
$\{y\}_{m_1(\{y\})}$	$\{\}_{m_1(\{y\})m_2(\{x\})}$	$\{y\}_{m_1(\{y\})m_2(\{y\})}$	$\{\}_{m_1(\{y\})m_2(\{z\})}$	$\{y\}_{m_1(\{y\})m_2(\Theta)}$
$\{z\}_{m_1(\{z\})}$	$\{\}_{m_1(\{z\})m_2(\{x\})}$	$\{\}_{m_1(\{z\})m_2(\{y\})}$	$\{z\}_{m_1(\{z\})m_2(\{z\})}$	$\{z\}_{m_1(\{z\})m_2(\Theta)}$
$\Theta_{m_1(\Theta)}$	$\{x\}_{m_1(\Theta)m_2(\{x\})}$	$\{y\}_{m_1(\Theta)m_2(\{y\})}$	$\{z\}_{m_1(\Theta)m_2(\{z\})}$	$\Theta_{m_1(\Theta)m_2(\Theta)}$

Note that if A_1, A_2, B_1, B_2 are singletons, where $A_2 = B_1$, the values of their belief functions are the same as the values of mass functions m_1 and m_2, respectively. Therefore, we have a set of strengths of belief with three possible categories as a combined result: $(bel(A_1), bel(A_2), bel(B_2))$.

On one hand, the mass function given by SVM is

$$m_1(A_1) = m_1(\{c_1\}) = 0.724, m_1(A_2) = m_1(\{c_2\}) = 0.184,$$

and the ignorance $m_1(C) = 0.092$. By choosing the category with the maximum degree of belief as a decision, it seems that document 37928 corresponds to $A_1 = \{c_1\}$—the decision made by the SVM classifier.

On the other hand, the mass function given by kNN is $m_2(B_1) = m_2(\{c_2\}) = 0.688, m_2(B_2) = m_2(\{c_4\}) = 0.208$, and the ignorance $m_2(C) = 0.104$. By choosing the category with the maximum degree of belief as a decision, we see that document 37928 corresponds to $B_1 = \{c_2\}$—the decision made by the kNN classifier.

By computing the pairwise orthogonal sum as in Fig. 2, we combine all of the 2-points focused mass functions. We can obtain a set of aggregated results as follows: First of all, we compute

$$\begin{aligned}&(m_1 \oplus m_2)(\{c_1\}) = 0.24, (m_1 \oplus m_2)(\{c_2\}) = 0.67,\\ &(m_1 \oplus m_2)(\{c_4\}) = 0.06, (m_1 \oplus m_2)(C) = 0.03.\end{aligned}$$

Then, we find the associated mass function

$$\begin{aligned}&m(\{c_2\}) = 0.67, m(\{c_1\}) = 0.24,\\ &m(C) \quad = 1 - 0.67 - 0.24 = 0.03 + 0.06 = 0.09.\end{aligned}$$

Thus, the final decision made by the combined classifier is category c_2.

One way that we could improve on this method is by using a threshold for the allocation to ignorance. For example, we could use a threshold of 0.1 for ignorance—i.e., the mass allocated to the whole set (Frame of Discernment) is 0.1. The order of the final choices might then be different, although this is unlikely in the example above. It would also be useful to know, in practice, conditions under which our triplet would be improved upon by using a *quartet*. That is, when would a 3-points focused mass function be better? To study this, we carry out a simple analysis for three categories. There are a limited number of permutations of categories A_1, A_2, A_3 to consider and, by the nature of the problem space, some of these cannot occur. For example, the two orders would not start with the same category as the most strongly supported category in both lists will then be the same and it would clearly be the overall winner. For clarity of presentation, we remove the ordering condition of Definition 3.

Consider the combination of two pieces of evidence m_1 and m_2 in the case where we have such a quartet rather than a triplet and the three sets are the same, but are not necessarily supported the same in each case. So, we keep the best three categories A_1, A_2, A_3, where $A_1 = \{x\}, A_2 = \{y\}, A_3 = \{z\}$. Using an intersection table (see Table 1), we get the orthogonal sum of two mass functions in the best three categories as shown in Table 1:

$$\begin{aligned}(m_1 \oplus m_2)(\{x\}) &= K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta)\\ &\quad + m_1(\Theta)m_2(\{x\})),\\ (m_1 \oplus m_2)(\{y\}) &= K(m_1(\{y\})m_2(\{y\}) + m_1(\{y\})m_2(\Theta)\\ &\quad + m_1(\Theta)m_2(\{y\})),\\ (m_1 \oplus m_2)(\{z\}) &= K(m_1(\{z\})m_2(\{z\}) + m_1(\{z\})m_2(\Theta)\\ &\quad + m_1(\Theta)m_2(\{z\})),\\ (m_1 \oplus m_2)(\Theta) &= Km_1(\Theta)m_2(\Theta),\end{aligned}$$

where

$$\begin{aligned}1/K = N = 1 &- m_1(\{x\})m_2(\{y\}) - m_1(\{y\})m_2(\{x\})\\ &- m_1(\{x\})m_2(\{z\}) - m_1(\{z\})m_2(\{x\})\\ &- m_1(\{y\})m_2(\{z\}) - m_1(\{z\})m_2(\{y\}) > 0.\end{aligned}$$

We can then draw some interesting conclusions for particular cases using some simple algebra. For example, we can rewrite the conditions for A_1 being best as:

Condition 1.

$$\begin{aligned}&m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) + m_1(\Theta)m_2(\{x\})\\ &\quad > m_1(\{y\})m_2(\{y\}) + m_1(\{y\})m_2(\Theta) + m_1(\Theta)m_2(\{y\}),\end{aligned}$$

i.e.,

$$\begin{aligned}&m_1(\{x\})[m_2(\{x\}) + m_2(\Theta)]\\ &> m_1(\{y\})m_2(\{y\}) + m_1(\{y\})m_2(\Theta)\\ &- m_1(\Theta)[m_2(\{x\}) - m_2(\{y\}).\end{aligned}$$

Condition 2.

$$m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) + m_1(\Theta)m_2(\{x\}) > m_1(\{z\})m_2(\{z\}) + m_1(\{z\})m_2(\Theta) + m_1(\Theta)m_2(\{z\}),$$

i.e.,

$$m_1(\{x\})[m_2(\{x\}) + m_2(\Theta)] > m_1(\{z\})m_2(\{z\}) + m_1(\{z\})m_2(\Theta) - m_1(\Theta)[m_2(\{z\}) - m_2(\{x\})].$$

Then, we can say such things as: When

$$m_1(\{x\})[m_2(\{x\}) + m_2(\Theta)] > m_1(\{y\})m_2(\{y\}),$$

the first condition *always* holds when $m_1(\Theta) > m_2(\Theta)$ and $[m_2(\{x\}) - m_2(\{y\})] > m_1(\{y\})$. This would occur when the first set of results has $m_1(\{x\}) \geq m_1(\{y\}) \geq m_1(\{z\})$ and the second set of results has $m_2(\{z\}) > m_2(\{x\}) > m_2(\{y\})$.

Suppose, for illustration, that the second set of results brings the third rated category to the top. That is, $m_2(\{z\}) > \max(m_2(\{x\}), m_2(\{y\}), m_2(\Theta))$.

This would happen, for example, if we had

$$m_1(\{x\}) = 0.6, m_1(\{y\}) = 0.2, m_1(\{z\}) = 0.1, m_1(\Theta) = 0.1;$$

and

$$m_2(\{z\}) = 0.45, m_2(\{x\}) = 0.35, \quad m_2(\{y\}) = 0.14, m_2(\Theta) = 0.06.$$

Now, for A_3 to be better supported than A_1, we need to have

$$m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) + m_1(\Theta)m_2(\{x\}) < m_1(\{z\})m_2(\{z\}) + m_1(\{z\})m_2(\Theta) + m_1(\Theta)m_2(\{z\}),$$

i.e.,

$$m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) + [1 - m_1(\{x\}) - m_1(\{y\}) - m_1(\{z\})]m_2(\{x\}) < m_1(\{z\})m_2(\{z\}) + m_1(\{z\})m_2(\Theta) + [1 - m_1(\{x\}) - m_1(\{y\}) - m_1(\{z\})]m_2(\{z\}),$$

i.e.,

$$m_1(\{x\})m_2(\Theta) + [1 - m_1(\{y\}) - m_1(\{z\})]m_2(\{x\}) < m_1(\{z\})m_2(\Theta) + [1 - m_1(\{x\}) - m_1(\{y\})]m_2(\{z\}),$$

i.e.,

$$m_1(\{x\})[1 - m_2(\{x\}) - m_2(\{y\}) - m_2(\{z\})] + [1 - m_1(\{y\}) - m_1(\{z\})]m_2(\{x\}) < m_1(\{z\})[1 - m_2(\{x\}) - m_2(\{y\}) - m_2(\{z\})] + [1 - m_1(\{x\}) - m_1(\{y\})]m_2(\{z\}),$$

i.e.,

$$m_1(\{x\})[1 - m_2(\{x\}) - m_2(\{y\})] + [1 - m_1(\{y\})]m_2(\{x\}) < m_1(\{z\})[1 - m_2(\{y\}) - m_2(\{z\})] + [1 - m_1(\{y\})]m_2(\{z\}),$$

i.e.,

$$m_1(\{x\})[1 - m_2(\{x\}) - m_2(\{y\})] < m_1(\{z\})[1 - m_2(\{y\}) - m_2(\{z\})] + [1 - m_1(\{y\})][m_2(\{z\}) - m_2(\{x\})],$$

i.e.,

$$m_1(\{x\}) < \frac{m_1(\{z\})[1 - m_2(\{y\}) - m_2(\{z\})]}{1 - m_2(\{x\}) - m_2(\{y\})} + \frac{[1 - m_1(\{y\})][m_2(\{z\}) - m_2(\{x\})]}{1 - m_2(\{x\}) - m_2(\{y\})}.$$

Now, consider the case when only the best two of each categorization method are used (i.e., using a triplet as in [8]).

Let $A_1 = \{x\}, A_2 = \{y\}, A_3 = \{z\}$, and let m_1, m_2 be two 2-points focused mass functions showing one equal point

$$m_1(\{x\}) + m_1(\{y\}) + m_1(\Theta) = 1; 0 \leq m_1(\{x\}), m_1(\{y\}), m_1(\Theta) \leq 1;$$
$$m_2(\{x\}) + m_2(\{z\}) + m_2(\Theta) = 1;\ 0 \leq m_2(\{x\}), m_2(\{z\}), m_2(\Theta) \leq 1.$$

To combine m_1, m_2, make the intersection table for $m_1 \oplus m_2$ as shown in Table 2. Then, we find that

TABLE 2
Intersection Table of Two Triplets

$m_1 \oplus m_2$	$\{x\}_{m_2(\{x\})}$	$\{z\}_{m_2(\{z\})}$	$\Theta_{m_2(\Theta)}$
$\{x\}_{m_1(\{x\})}$	$\{x\}_{m_1(\{x\})m_2(\{x\})}$	$\{\}_{m_1(\{x\})m_2(\{z\})}$	$\{x\}_{m_1(\{x\})m_2(\Theta)}$
$\{y\}_{m_1(\{y\})}$	$\{\}_{m_1(\{y\})m_2(\{x\})}$	$\{\}_{m_1(\{y\})m_2(\{z\})}$	$\{y\}_{m_1(\{y\})m_2(\Theta)}$
$\Theta_{m_1(\Theta)}$	$\{x\}_{m_1(\Theta)m_2(\{x\})}$	$\{z\}_{m_1(\Theta)m_2(\{z\})}$	$\Theta_{m_1(\Theta)m_2(\Theta)}$

$$(m_1 \oplus m_2)(\{x\}) = K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\}) m_2(\Theta) + m_1(\Theta)m_2(\{x\})),$$
$$(m_1 \oplus m_2)(\{y\}) = Km_1(\{y\})m_2(\Theta),$$
$$(m_1 \oplus m_2)(\{z\}) = Km_1(\Theta)m_2(\{z\}),$$
$$(m_1 \oplus m_2)(\Theta) \ = Km_1(\Theta)m_2(\Theta),$$

where

$$1/K = 1 - m_1(\{x\})m_2(\{z\}) - m_1(\{y\})m_2(\{z\}) - m_1(\{y\}) m_2(\{x\}) > 0.$$

A_1 is the better choice when:

$$m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) + m_1(\Theta)m_2(\{x\}) > m_1(\Theta)m_2(\{z\}),$$

i.e.,

$$[m_1(\{x\}) + m_1(\Theta)]m_2(\{x\}) + m_1(\{x\})[1 - m_2(\{x\}) - m_2(\{z\})] > [1 - m_1(\{x\}) - m_1(\{y\})]m_2(\{z\}),$$

i.e.,

$$[1 - m_1(\{y\})]m_2(\{x\}) + m_1(\{x\})[1 - m_2(\{x\})] > [1 - m_1(\{y\})]m_2(\{z\}),$$

i.e.,

$$m_1(\{x\})[1 - m_2(\{x\})] > [1 - m_1(\{y\})][m_2(\{z\}) - m_2(\{x\})],$$

i.e.,

$$m_1(\{x\}) > \frac{[1 - m_1(\{y\})][m_2(\{z\}) - m_2(\{x\})]}{1 - m_2(\{x\})}.$$

For example, if $m_2(\{x\}) = 0.1$ and $m_1(\{y\}) = 0.1$ or more, then $m_1(\{x\}) > m_2(\{z\}) - 0.1$.

Direct Comparison. When $m_2(\{y\})$ is small, $m_1(\{x\}) > \frac{[1-m_1(\{y\})][m_2(\{z\})-m_2(\{x\})]}{1-m_2(\{x\})}$ means A_1 is best for a triplet.

$$m_1(\{x\}) < \frac{m_1(\{z\})[1 - m_2(\{y\}) - m_2(\{z\})]}{1 - m_2(\{x\}) - m_2(\{y\})} + \frac{[1 - m_1(\{y\})][m_2(\{z\}) - m_2(\{x\})]}{1 - m_2(\{x\}) - m_2(\{y\})}$$

means A_3 is better for a quartet.

So, considering two foci gives A_1 *better* when $m_1(\{x\}) > \frac{[1-m_1(\{y\})][m_2(\{z\})-m_2(\{x\})]}{1-m_2(\{x\})}$.

However, when a quartet, rather than a triplet ,is considered, this changes to A_3 *better* when $m_1(\{z\})$ is such that

$$m_1(\{x\}) < \frac{m_1(\{z\})[1 - m_2(\{y\}) - m_2(\{z\})]}{1 - m_2(\{x\}) - m_2(\{y\})} + \frac{[1 - m_1(\{y\})][m_2(\{z\}) - m_2(\{x\})]}{1 - m_2(\{x\}) - m_2(\{y\})}.$$

5 Constricted Mass Functions

To recapitulate, a mass function m is called *2-points focused* if it has no focuses other than two singletons and Θ. That is, there exist two elements $x, y \in \Theta$ such that

$$m(\{x\}) + m(\{y\}) + m(\Theta) = 1;$$
$$0 \le m(\{x\}), m(\{y\}), m(\Theta) \le 1.$$

Similarly, we can consider 3-points focused, 4-points focused, ..., n-points focused mass functions. We have discussed 2-points focused mass functions in other papers [8], and discuss 3-points more fully in this paper.

Generally, a mass function may have more than three focal singletons. We can use a *focusing operator* σ to a *constricted* mass function of 3-points as follows: Let m be a mass function with focal singletons $\{x_1\}, \{x_2\}, \ldots, \{x_n\}; n \ge 3$. Then, the focusing operator σ makes m^σ as:

$$m^\sigma(\{u\}) + m^\sigma(\{v\}) + m^\sigma(\{w\}) + m^\sigma(\Theta) = 1,$$

where $u = x_{i_1}, v = x_{i_2}, w = x_{i_3}$ for

$$m(\{x_{i_1}\}) \ge m(\{x_{i_2}\}) \ge m(\{x_{i_3}\}) \ge \ldots \ge m(\{x_{i_n}\})$$

and $m^\sigma(\Theta) = 1 - m^\sigma(\{u\}) - m^\sigma(\{v\}) - m^\sigma(\{w\})$.

We seek general formulae for combining 3-points focused evidential functions. These give the basis of our combination algorithm.

5.1 Three Points Equal

We now consider again the case where the three foci are equal (e.g., x, y, z in our example).

Theorem 5.1.1. *Let m_1, m_2 be two 3-points focused mass functions having three points equal,*

$$m_1(\{x\}) + m_1(\{y\}) + m_1(\{z\}) + m_1(\Theta) = 1;$$
$$0 \le m_1(\{x\}), m_1(\{y\}), m_1(\{z\}), m_1(\Theta) \le 1;$$
$$m_2(\{x\}) + m_2(\{y\}) + m_2(\{z\}) + m_2(\Theta) = 1;$$
$$0 \le m_2(\{x\}), m_2(\{y\}), m_2(\{z\}), m_2(\Theta) \le 1.$$

Then,

$$\begin{aligned}1/K = N = 1 &- m_1(\{x\})m_2(\{y\}) - m_1(\{y\})m_2(\{x\}) \\ &- m_1(\{x\})m_2(\{z\}) - m_1(\{z\})m_2(\{x\}) \\ &- m_1(\{y\})m_2(\{z\}) - m_1(\{z\})m_2(\{y\}),\end{aligned}$$

and m_1, m_2 are combinable if and only if

$$m_1(\{x\})m_2(\{y\}) + m_1(\{y\})m_2(\{x\}) + m_1(\{x\})m_2(\{z\}) + m_1(\{z\})m_2(\{x\}) + m_1(\{y\})m_2(\{z\}) + m_1(\{z\})m_2(\{y\}) < 1.$$

When m_1, m_2 are combinable, we have

$$\begin{aligned}(m_1 \oplus m_2)(\{x\}) &= K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) + m_1(\Theta)m_2(\{x\})),\\ (m_1 \oplus m_2)(\{y\}) &= K(m_1(\{y\})m_2(\{y\}) + m_1(\{y\})m_2(\Theta) + m_1(\Theta)m_2(\{y\})),\\ (m_1 \oplus m_2)(\{z\}) &= K(m_1(\{z\})m_2(\{z\}) + m_1(\{z\})m_2(\Theta) + m_1(\Theta)m_2(\{z\})),\\ (m_1 \oplus m_2)(\Theta) &= Km_1(\Theta)m_2(\Theta).\end{aligned}$$

Proof. To combine m_1, m_2, make the intersection table for $m_1 \oplus m_2$ as shown in Table 3. Then, we find that

$$\begin{aligned}1/K = N = 1 - \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y) = 1 &- m_1(\{x\})m_2(\{y\}) \\ &- m_1(\{y\})m_2(\{x\}) - m_1(\{x\})m_2(\{z\}) \\ &- m_1(\{z\})m_2(\{x\}) - m_1(\{y\})m_2(\{z\}) \\ &- m_1(\{z\})m_2(\{y\}),\end{aligned}$$

and that m_1, m_2 are combinable if and only if $N > 0$, i.e.,

$$\begin{aligned}1 &- m_1(\{x\})m_2(\{y\}) - m_1(\{y\})m_2(\{x\}) - m_1(\{x\})m_2(\{z\}) \\ &- m_1(\{z\})m_2(\{x\}) - m_1(\{y\})m_2(\{z\}) \\ &- m_1(\{z\})m_2(\{y\}) > 0,\end{aligned}$$

i.e.,

$$\begin{aligned}m_1(\{x\})m_2(\{y\}) &+ m_1(\{y\})m_2(\{x\}) + m_1(\{x\})m_2(\{z\}) \\ &+ m_1(\{z\})m_2(\{x\}) + m_1(\{y\})m_2(\{z\}) \\ &+ m_1(\{z\})m_2(\{y\}) < 1.\end{aligned}$$

We know that

$$(m_1 \oplus m_2)(A) = K \sum_{X \cap Y = A} m_1(X)m_2(Y).$$

TABLE 3
Three Points Equal

$m_1 \oplus m_2$	$\{x\}_{m_2(\{x\})}$	$\{y\}_{m_2(\{y\})}$	$\{z\}_{m_2(\{z\})}$	$\Theta_{m_2(\Theta)}$
$\{x\}_{m_1(\{x\})}$	$\{x\}_{m_1(\{x\})m_2(\{x\})}$	$\{\}_{m_1(\{x\})m_2(\{y\})}$	$\{\}_{m_1(\{x\})m_2(\{z\})}$	$\{x\}_{m_1(\{x\})m_2(\Theta)}$
$\{y\}_{m_1(\{y\})}$	$\{\}_{m_1(\{y\})m_2(\{x\})}$	$\{y\}_{m_1(\{y\})m_2(\{y\})}$	$\{\}_{m_1(\{y\})m_2(\{z\})}$	$\{y\}_{m_1(\{y\})m_2(\Theta)}$
$\{z\}_{m_1(\{z\})}$	$\{\}_{m_1(\{z\})m_2(\{x\})}$	$\{\}_{m_1(\{z\})m_2(\{y\})}$	$\{z\}_{m_1(\{z\})m_2(\{z\})}$	$\{z\}_{m_1(\{z\})m_2(\Theta)}$
$\Theta_{m_1(\Theta)}$	$\{x\}_{m_1(\Theta)m_2(\{x\})}$	$\{y\}_{m_1(\Theta)m_2(\{y\})}$	$\{z\}_{m_1(\Theta)m_2(\{z\})}$	$\Theta_{m_1(\Theta)m_2(\Theta)}$

TABLE 4
Two Points Equal

$m_1 \oplus m_2$	$\{x\}_{m_2(\{x\})}$	$\{y\}_{m_2(\{y\})}$	$\{u\}_{m_2(\{u\})}$	$\Theta_{m_2(\Theta)}$
$\{x\}_{m_1(\{x\})}$	$\{x\}_{m_1(\{x\})m_2(\{x\})}$	$\{\}_{m_1(\{x\})m_2(\{y\})}$	$\{\}_{m_1(\{x\})m_2(\{u\})}$	$\{x\}_{m_1(\{x\})m_2(\Theta)}$
$\{y\}_{m_1(\{y\})}$	$\{\}_{m_1(\{y\})m_2(\{x\})}$	$\{y\}_{m_1(\{y\})m_2(\{y\})}$	$\{\}_{m_1(\{y\})m_2(\{u\})}$	$\{y\}_{m_1(\{y\})m_2(\Theta)}$
$\{z\}_{m_1(\{z\})}$	$\{\}_{m_1(\{z\})m_2(\{x\})}$	$\{\}_{m_1(\{z\})m_2(\{y\})}$	$\{\}_{m_1(\{z\})m_2(\{u\})}$	$\{z\}_{m_1(\{z\})m_2(\Theta)}$
$\Theta_{m_1(\Theta)}$	$\{x\}_{m_1(\Theta)m_2(\{x\})}$	$\{y\}_{m_1(\Theta)m_2(\{y\})}$	$\{u\}_{m_1(\Theta)m_2(\{u\})}$	$\Theta_{m_1(\Theta)m_2(\Theta)}$

By the intersection table for $m_1 \oplus m_2$ we find that

$$
\begin{aligned}
(m_1 \oplus m_2)(\{x\}) &= K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) \\
&\quad + m_1(\Theta)m_2(\{x\})), \\
(m_1 \oplus m_2)(\{y\}) &= K(m_1(\{y\})m_2(\{y\}) + m_1(\{y\})m_2(\Theta) \\
&\quad + m_1(\Theta)m_2(\{y\})), \\
(m_1 \oplus m_2)(\{z\}) &= K(m_1(\{z\})m_2(\{z\}) + m_1(\{z\})m_2(\Theta) \\
&\quad + m_1(\Theta)m_2(\{z\})), \\
(m_1 \oplus m_2)(\Theta) &= Km_1(\Theta)m_2(\Theta).
\end{aligned}
$$

□

5.2 Two Points Equal

Now, the three points equal case is probably the simplest. We need to consider other possibilities.

Theorem 5.2.1. *Let m_1, m_2 be two 3-points focused mass functions having points equal 2,*

$$
\begin{aligned}
&m_1(\{x\}) + m_1(\{y\}) + m_1(\{z\}) + m_1(\Theta) = 1; \\
&\qquad 0 \leq m_1(\{x\}), m_1(\{y\}), m_1(\{z\}), m_1(\Theta) \leq 1; \\
&m_2(\{x\}) + m_2(\{y\}) + m_2(\{u\}) + m_2(\Theta) = 1; \\
&\qquad 0 \leq m_2(\{x\}), m_2(\{y\}), m_2(\{u\}), m_2(\Theta) \leq 1.
\end{aligned}
$$

Then,

$$
\begin{aligned}
1/K = N = 1 &- m_1(\{x\})m_2(\{y\}) - m_1(\{y\})m_2(\{x\}) \\
&- m_1(\{z\})m_2(\{u\}) - m_1(\{x\})m_2(\{u\}) \\
&- m_1(\{z\})m_2(\{x\}) - m_1(\{y\})m_2(\{u\}) \\
&- m_1(\{z\})m_2(\{y\}),
\end{aligned}
$$

and m_1, m_2 are combinable if and only if

$$
\begin{aligned}
&m_1(\{x\})m_2(\{y\}) + m_1(\{y\})m_2(\{x\}) + m_1(\{x\})m_2(\{u\}) \\
&\quad + m_1(\{z\})m_2(\{x\}) + m_1(\{y\})m_2(\{u\}) \\
&\quad + m_1(\{z\})m_2(\{y\}) + m_1(\{z\})m_2(\{u\}) < 1.
\end{aligned}
$$

When m_1, m_2 are combinable we have

$$
\begin{aligned}
&(m_1 \oplus m_2)(\{x\}) = K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) \\
&\quad + m_1(\Theta)m_2(\{x\})), \\
&(m_1 \oplus m_2)(\{y\}) = K(m_1(\{y\})m_2(\{y\}) + m_1(\{y\})m_2(\Theta) \\
&\quad + m_1(\Theta)m_2(\{y\})), \\
&(m_1 \oplus m_2)(\{z\}) = Km_1(\{z\})m_2(\Theta), (m_1 \oplus m_2)(\{u\}) \\
&\quad = Km_1(\Theta)m_2(\{u\}), \\
&(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta).
\end{aligned}
$$

Proof. To combine m_1, m_2, make the intersection table for $m_1 \oplus m_2$ as shown in Table 4. Then, we find that

$$
\begin{aligned}
1/K = N &= 1 - \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y) \\
&= 1 - m_1(\{x\})m_2(\{y\}) - m_1(\{y\})m_2(\{x\}) \\
&\quad - m_1(\{z\})m_2(\{u\}) - m_1(\{x\})m_2(\{u\}) \\
&\quad - m_1(\{z\})m_2(\{x\}) - m_1(\{y\})m_2(\{u\}) \\
&\quad - m_1(\{z\})m_2(\{y\}),
\end{aligned}
$$

and that m_1, m_2 are combinable if and only if $N > 0$, i.e.,

$$
\begin{aligned}
&1 - m_1(\{x\})m_2(\{y\}) - m_1(\{y\})m_2(\{x\}) - m_1(\{z\})m_2(\{u\}) \\
&- m_1(\{x\})m_2(\{u\}) - m_1(\{z\})m_2(\{x\}) - m_1(\{y\})m_2(\{u\}) \\
&- m_1(\{z\})m_2(\{y\}) > 0,
\end{aligned}
$$

i.e.,

$$m_1(\{x\})m_2(\{y\}) + m_1(\{y\})m_2(\{x\}) + m_1(\{x\})m_2(\{u\}) + m_1(\{z\})m_2(\{x\}) + m_1(\{y\}) m_2(\{u\}) + m_1(\{z\})m_2(\{y\}) + m_1(\{z\})m_2(\{u\}) < 1.$$

We know that

$$(m_1 \oplus m_2)(A) = K \sum_{X \cap Y = A} m_1(X)m_2(Y).$$

By the intersection table for $m_1 \oplus m_2$ we find that

$$(m_1 \oplus m_2)(\{x\}) = K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) + m_1(\Theta)m_2(\{x\})),$$
$$(m_1 \oplus m_2)(\{y\}) = K(m_1(\{y\})m_2(\{y\}) + m_1(\{y\})m_2(\Theta) + m_1(\Theta)m_2(\{y\})),$$
$$(m_1 \oplus m_2)(\{z\}) = Km_1(\{z\})m_2(\Theta),$$
$$(m_1 \oplus m_2)(\{u\}) = Km_1(\Theta)m_2(\{u\}),$$
$$(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta).$$

□

However, now that $m_1 \oplus m_2$ is not a 3-points focused mass function, there are focal points $\{x\}, \{y\}, \{z\}, \{u\}$, and Θ. The 3-points focusing operator σ should be applied.

Theorem 5.2.2. *Let m_1, m_2 be two 3-points focused mass functions,*

$$m_1(\{x\}) + m_1(\{y\}) + m_1(\{z\}) + m_1(\Theta) = 1;$$
$$0 \leq m_1(\{x\}), m_1(\{y\}), m_1(\{z\}), m_1(\Theta) \leq 1;$$
$$m_2(\{x\}) + m_2(\{y\}) + m_2(\{u\}) + m_2(\Theta) = 1;$$
$$0 \leq m_2(\{x\}), m_2(\{y\}), m_2(\{u\}), m_2(\Theta) \leq 1.$$

Suppose that

$$m_1(\{x\})m_2(\{y\}) + m_1(\{y\})m_2(\{x\}) + m_1(\{x\})m_2(\{u\}) + m_1(\{z\})m_2(\{x\}) + m_1(\{y\})m_2(\{u\}) + m_1(\{z\})m_2(\{y\}) + m_1(\{z\})m_2(\{u\}) < 1.$$

So, m_1, m_2 are combinable. Denote

$$(m_1 \oplus m_2)(\{x\}) = K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) + m_1(\Theta)m_2(\{x\})) = f(x),$$
$$(m_1 \oplus m_2)(\{y\}) = K(m_1(\{y\})m_2(\{y\}) + m_1(\{y\})m_2(\Theta) + m_1(\Theta)m_2(\{y\})) = f(y),$$
$$(m_1 \oplus m_2)(\{z\}) = Km_1(\{z\})m_2(\Theta) = f(z),$$
$$(m_1 \oplus m_2)(\{u\}) = Km_1(\Theta)m_2(\{u\}) = f(u),$$
$$(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta),$$

where

$$1/K = N = 1 - m_1(\{x\})m_2(\{y\}) - m_1(\{y\})m_2(\{x\}) - m_1(\{z\})m_2(\{u\}) - m_1(\{x\})m_2(\{u\}) - m_1(\{z\})m_2(\{x\}) - m_1(\{y\})m_2(\{u\}) - m_1(\{z\})m_2(\{y\}).$$

Then, $(m_1 \oplus m_2)^\sigma$ is the following:

$$(m_1 \oplus m_2)^\sigma(\{x'\}) + (m_1 \oplus m_2)^\sigma(\{y'\}) + (m_1 \oplus m_2)^\sigma(\{z'\}) + (m_1 \oplus m_2)^\sigma(\Theta) = 1, (m_1 \oplus m_2)^\sigma(\{x'\}) = f(x'),$$

where $\{x', y', z', u'\} = \{x, y, z, u\}$ and

$$f(x') \geq f(y') \geq f(z') \geq f(u'),$$

and

$$(m_1 \oplus m_2)^\sigma(\{y'\}) = f(y'),$$
$$(m_1 \oplus m_2)^\sigma(\{z'\}) = f(z'),$$
$$(m_1 \oplus m_2)^\sigma(\Theta) = 1 - f(x') - f(y') - f(z').$$

Proof. By Theorem 5.2.1,

$$(m_1 \oplus m_2)(\{x\}) = K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) + m_1(\Theta)m_2(\{x\})) = f(x),$$
$$(m_1 \oplus m_2)(\{y\}) = K(m_1(\{y\})m_2(\{y\}) + m_1(\{y\})m_2(\Theta) + m_1(\Theta)m_2(\{y\})) = f(y),$$
$$(m_1 \oplus m_2)(\{z\}) = Km_1(\{z\})m_2(\Theta) = f(z),$$
$$(m_1 \oplus m_2)(\{u\}) = Km_1(\Theta)m_2(\{u\}) = f(u),$$
$$(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta),$$

where

$$1/K = N = 1 - m_1(\{x\})m_2(\{y\}) - m_1(\{y\})m_2(\{x\}) - m_1(\{z\})m_2(\{u\}) - m_1(\{x\})m_2(\{u\}) - m_1(\{z\})m_2(\{x\}) - m_1(\{y\})m_2(\{u\}) - m_1(\{z\})m_2(\{y\}).$$

Then, by the definition of the 3-points focusing operator σ, we find that $(m_1 \oplus m_2)^\sigma$ as follows:

$$(m_1 \oplus m_2)^\sigma(\{x'\}) + (m_1 \oplus m_2)^\sigma(\{y'\}) + (m_1 \oplus m_2)^\sigma(\{z'\}) + (m_1 \oplus m_2)^\sigma(\Theta) = 1, (m_1 \oplus m_2)^\sigma(\{x'\}) = f(x'),$$

where $\{x', y', z', u'\} = \{x, y, z, u\}$ and

$$f(x') \geq f(y') \geq f(z') \geq f(u'),$$

and $(m_1 \oplus m_2)^\sigma(\{y'\}) = f(y')$, $(m_1 \oplus m_2)^\sigma(\{z'\}) = f(z')$, $(m_1 \oplus m_2)^\sigma(\Theta) = 1 - f(x') - f(y') - f(z')$. □

5.3 One Equal Point

Only one focus might be shared by the two sets of results.

Theorem 5.3.1. *Let m_1, m_2 be two 3-points focused mass functions having one equal point,*

$$m_1(\{x\}) + m_1(\{y\}) + m_1(\{z\}) + m_1(\Theta) = 1;$$
$$0 \leq m_1(\{x\}), m_1(\{y\}), m_1(\{z\}), m_1(\Theta) \leq 1;$$
$$m_2(\{x\}) + m_2(\{u\}) + m_2(\{v\}) + m_2(\Theta) = 1;$$
$$0 \leq m_2(\{x\}), m_2(\{u\}), m_2(\{v\}), m_2(\Theta) \leq 1.$$

Then,

$$1/K = N = 1 - m_1(\{x\})m_2(\{u\}) - m_1(\{y\})m_2(\{x\}) - m_1(\{z\})m_2(\{v\} - m_1(\{y\})m_2(\{u\}) - m_1(\{x\})m_2(\{v\}) - m_1(\{z\})m_2(\{x\}) - m_1(\{y\})m_2(\{v\}) - m_1(\{z\})m_2(\{u\}),$$

TABLE 5
One Equal Point

$m_1 \oplus m_2$	$\{x\}_{m_2(\{x\})}$	$\{u\}_{m_2(\{u\})}$	$\{v\}_{m_2(\{v\})}$	$\Theta_{m_2(\Theta)}$
$\{x\}_{m_1(\{x\})}$	$\{x\}_{m_1(\{x\})m_2(\{x\})}$	$\{\}_{m_1(\{x\})m_2(\{u\})}$	$\{\}_{m_1(\{x\})m_2(\{v\})}$	$\{x\}_{m_1(\{x\})m_2(\Theta)}$
$\{y\}_{m_1(\{y\})}$	$\{\}_{m_1(\{y\})m_2(\{x\})}$	$\{\}_{m_1(\{y\})m_2(\{u\})}$	$\{\}_{m_1(\{y\})m_2(\{v\})}$	$\{y\}_{m_1(\{y\})m_2(\Theta)}$
$\{z\}_{m_1(\{z\})}$	$\{\}_{m_1(\{z\})m_2(\{x\})}$	$\{\}_{m_1(\{z\})m_2(\{u\})}$	$\{\}_{m_1(\{z\})m_2(\{v\})}$	$\{z\}_{m_1(\{z\})m_2(\Theta)}$
$\Theta_{m_1(\Theta)}$	$\{x\}_{m_1(\Theta)m_2(\{x\})}$	$\{u\}_{m_1(\Theta)m_2(\{u\})}$	$\{v\}_{m_1(\Theta)m_2(\{v\})}$	$\Theta_{m_1(\Theta)m_2(\Theta)}$

and m_1, m_2 are combinable if and only if

$$\begin{aligned} &m_1(\{x\})m_2(\{u\}) + m_1(\{y\})m_2(\{x\}) + m_1(\{x\})m_2(\{v\}) \\ &\quad + m_1(\{z\})m_2(\{x\}) + m_1(\{y\})m_2(\{v\}) + m_1(\{z\})m_2(\{u\}) \\ &\quad + m_1(\{z\})m_2(\{v\} + m_1(\{y\})m_2(\{u\}) < 1. \end{aligned}$$

When m_1, m_2 are combinable we have

$$\begin{aligned} &(m_1 \oplus m_2)(\{x\}) = K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) \\ &\quad + m_1(\Theta)m_2(\{x\})), \\ &(m_1 \oplus m_2)(\{y\}) = Km_1(\{y\})m_2(\Theta), \\ &(m_1 \oplus m_2)(\{z\}) = Km_1(\{z\})m_2(\Theta), \\ &(m_1 \oplus m_2)(\{u\}) = Km_1(\Theta)m_2(\{u\}), \\ &(m_1 \oplus m_2)(\{v\}) = Km_1(\Theta)m_2(\{v\}), \\ &(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta), \end{aligned}$$

Proof. To combine m_1, m_2, make the intersection table for $m_1 \oplus m_2$ as shown in Table 5. Then, we find that

$$\begin{aligned} 1/K = N = 1 - &\sum_{X \cap Y = \emptyset} m_1(X)m_2(Y) = 1 - m_1(\{x\})m_2(\{u\}) \\ &- m_1(\{y\})m_2(\{x\}) - m_1(\{z\})m_2(\{v\}1 - m_1(\{y\})m_2(\{u\}) \\ &- m_1(\{x\})m_2(\{v\}) - m_1(\{z\})m_2(\{x\}) - m_1(\{y\})m_2(\{v\}) \\ &- m_1(\{z\})m_2(\{u\}), \end{aligned}$$

and that m_1, m_2 are combinable if and only if $N > 0$, i.e.,

$$\begin{aligned} &1 - m_1(\{x\})m_2(\{u\}) - m_1(\{y\})m_2(\{x\}) - m_1(\{z\})m_2(\{v\} \\ &\quad - m_1(\{y\})m_2(\{u\}) - m_1(\{x\})m_2(\{v\}) - m_1(\{z\})m_2(\{x\}) \\ &\quad - m_1(\{y\})m_2(\{v\}) - m_1(\{z\})m_2(\{u\}) > 0, \end{aligned}$$

i.e.,

$$\begin{aligned} &m_1(\{x\})m_2(\{u\}) + m_1(\{y\})m_2(\{x\}) + m_1(\{x\})m_2(\{v\}) \\ &\quad + m_1(\{z\})m_2(\{x\}) + m_1(\{y\})m_2(\{v\}) + m_1(\{z\})m_2(\{u\}) \\ &\quad + m_1(\{z\})m_2(\{v\} + m_1(\{y\})m_2(\{u\}) < 1. \end{aligned}$$

We know that

$$(m_1 \oplus m_2)(A) = K \sum_{X \cap Y = A} m_1(X)m_2(Y).$$

By the intersection table for $m_1 \oplus m_2$, we find that

$$\begin{aligned} &(m_1 \oplus m_2)(\{x\}) = K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) \\ &\quad + m_1(\Theta)m_2(\{x\})), \\ &(m_1 \oplus m_2)(\{y\}) = Km_1(\{y\})m_2(\Theta), (m_1 \oplus m_2)(\{z\}) \\ &\quad = Km_1(\{z\})m_2(\Theta), \\ &(m_1 \oplus m_2)(\{u\}) = Km_1(\Theta)m_2(\{u\}), \\ &(m_1 \oplus m_2)(\{v\}) = Km_1(\Theta)m_2(\{v\}), \\ &(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta). \end{aligned}$$

However, now $m_1 \oplus m_2$ is not a 3-points focused mass function, there are focal points $\{x\}, \{y\}, \{z\}, \{u\}, \{v\}$ and Θ. The 3-points focusing operator σ should again be applied. □

Theorem 5.3.2. *Let m_1, m_2 be two 3-points focused mass functions,*

$$\begin{aligned} &m_1(\{x\}) + m_1(\{y\}) + m_1(\{z\}) + m_1(\Theta) = 1;\ 0 \le m_1(\{x\}), \\ &m_1(\{y\}), m_1(\{z\}), m_1(\Theta) \le 1; \\ &m_2(\{x\}) + m_2(\{y\}) + m_2(\{u\}) + m_2(\Theta) = 1; \\ &0 \le m_2(\{x\}), m_2(\{y\}), m_2(\{u\}), m_2(\Theta) \le 1. \end{aligned}$$

Suppose that

$$\begin{aligned} &m_1(\{x\})m_2(\{u\}) + m_1(\{y\})m_2(\{x\}) + m_1(\{x\})m_2(\{v\}) \\ &\quad + m_1(\{z\})m_2(\{x\}) + m_1(\{y\})m_2(\{v\}) + m_1(\{z\})m_2(\{u\}) \\ &\quad + m_1(\{z\})m_2(\{v\} + m_1(\{y\})m_2(\{u\}) < 1. \end{aligned}$$

So, m_1, m_2 are combinable. Denote

$$\begin{aligned} &(m_1 \oplus m_2)(\{x\}) = K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta) \\ &\quad + m_1(\Theta)m_2(\{x\})) = f(x), \\ &(m_1 \oplus m_2)(\{y\}) = K(m_1(\{y\})m_2(\{y\}) + m_1(\{y\})m_2(\Theta) \\ &\quad + m_1(\Theta)m_2(\{y\})) = f(y), \\ &(m_1 \oplus m_2)(\{z\}) = Km_1(\{z\})m_2(\Theta) = f(z), \\ &(m_1 \oplus m_2)(\{u\}) = Km_1(\Theta)m_2(\{u\}) = f(u), \\ &(m_1 \oplus m_2)(\{v\}) = Km_1(\Theta)m_2(\{u\}) = f(v), \\ &(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta), \end{aligned}$$

where

$$\begin{aligned}1/K = N = 1 &- m_1(\{x\})m_2(\{u\}) - m_1(\{y\})m_2(\{x\})\\ &- m_1(\{z\})m_2(\{v\} - m_1(\{y\})m_2(\{u\})\\ &- m_1(\{x\})m_2(\{v\}) - m_1(\{z\})m_2(\{x\})\\ &- m_1(\{y\})m_2(\{v\}) - m_1(\{z\})m_2(\{u\}).\end{aligned}$$

Then, $(m_1 \oplus m_2)^\sigma$ *is the following:*

$$\begin{aligned}&(m_1 \oplus m_2)^\sigma(\{x'\}) + (m_1 \oplus m_2)^\sigma(\{y'\}) + (m_1 \oplus m_2)^\sigma(\{z'\})\\ &\quad + (m_1 \oplus m_2)^\sigma(\Theta) = 1,\\ &(m_1 \oplus m_2)^\sigma(\{x'\}) = f(x'),\end{aligned}$$

where $\{x', y', z', u', v'\} = \{x, y, z, u, v\}$ *and*

$$f(x') \geq f(y') \geq f(z') \geq f(u') \geq f(v'),$$

and $(m_1 \oplus m_2)^\sigma(\{y'\}) = f(y')$, $(m_1 \oplus m_2)^\sigma(\{z'\}) = f(z')$, $(m_1 \oplus m_2)^\sigma(\Theta) = 1 - f(x') - f(y') - f(z')$.

Proof. By Theorem 5.3.1,

$$\begin{aligned}&(m_1 \oplus m_2)(\{x\}) = K(m_1(\{x\})m_2(\{x\}) + m_1(\{x\})m_2(\Theta)\\ &\quad + m_1(\Theta)m_2(\{x\})) = f(x),\\ &(m_1 \oplus m_2)(\{y\}) = Km_1(\{y\})m_2(\Theta) = f(y), (m_1 \oplus m_2)(\{z\})\\ &\quad = Km_1(\{z\})m_2(\Theta) = f(z),\\ &(m_1 \oplus m_2)(\{u\}) = Km_1(\Theta)m_2(\{u\}) = f(u),\\ &(m_1 \oplus m_2)(\{v\}) = Km_1(\Theta)m_2(\{v\}) = f(v),\\ &(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta),\end{aligned}$$

where

$$\begin{aligned}1/K = N = 1 &- m_1(\{x\})m_2(\{u\}) - m_1(\{y\})m_2(\{x\})\\ &- m_1(\{z\})m_2(\{v\} - m_1(\{y\})m_2(\{u\})\\ &- m_1(\{x\})m_2(\{v\}) - m_1(\{z\})m_2(\{x\})\\ &- m_1(\{y\})m_2(\{v\}) - m_1(\{z\})m_2(\{u\}).\end{aligned}$$

Then, by the definition of the 3-points focusing operator σ we find that $(m_1 \oplus m_2)^\sigma$ as follows:

$$\begin{aligned}&(m_1 \oplus m_2)^\sigma(\{x'\}) + (m_1 \oplus m_2)^\sigma(\{y'\}) + (m_1 \oplus m_2)^\sigma(\{z'\})\\ &\quad + (m_1 \oplus m_2)^\sigma(\Theta) = 1,\\ &(m_1 \oplus m_2)^\sigma(\{x'\}) = f(x'),\end{aligned}$$

where $\{x', y', z', u', v'\} = \{x, y, z, u, v\}$ and

$$f(x') \geq f(y') \geq f(z') \geq f(u') \geq f(v'),$$

and $(m_1 \oplus m_2)^\sigma(\{y'\}) = f(y')$, $(m_1 \oplus m_2)^\sigma(\{z'\}) = f(z')$, $(m_1 \oplus m_2)^\sigma(\Theta) = 1 - f(x') - f(y') - f(z')$. □

5.4 Totally Different Points

Of course, there is no guarantee that *any* focus is shared between the two sets of results.

Theorem 5.4.1. *Let* m_1, m_2 *be two 3-points focused mass functions having one equal point,*

$$\begin{aligned}&m_1(\{x\}) + m_1(\{y\}) + m_1(\{z\}) + m_1(\Theta) = 1;\\ &0 \leq m_1(\{x\}), m_1(\{y\}), m_1(\{z\}), m_1(\Theta) \leq 1;\\ &m_2(\{u\}) + m_2(\{v\}) + m_2(\{w\}) + m_2(\Theta) = 1;\\ &0 \leq m_2(\{u\}), m_2(\{v\}), m_2(\{w\}), m_2(\Theta) \leq 1.\end{aligned}$$

Then,

$$\begin{aligned}1/K = N = 1 &- m_1(\{x\})m_2(\{u\}) - m_1(\{y\})m_2(\{u\})\\ &- m_1(\{z\})m_2(\{u\} - m_1(\{x\})m_2(\{v\})\\ &- m_1(\{y\})m_2(\{v\}) - m_1(\{z\})m_2(\{v\}\\ &- m_1(\{x\})m_2(\{w\}) - m_1(\{y\})m_2(\{w\})\\ &- m_1(\{z\})m_2(\{w\},\end{aligned}$$

and m_1, m_2 *are combinable if and only if*

$$\begin{aligned}&m_1(\{x\})m_2(\{u\}) + m_1(\{y\})m_2(\{u\}) + m_1(\{z\})m_2(\{u\})\\ &\quad + m_1(\{x\})m_2(\{v\}) + m_1(\{y\})m_2(\{v\}) + m_1(\{z\})m_2(\{v\})\\ &\quad + m_1(\{x\})m_2(\{w\}) + m_1(\{y\})m_2(\{w\}) + m_1(\{z\})m_2(\{w\})\\ &\quad < 1.\end{aligned}$$

When m_1, m_2 *are combinable we have*

$$\begin{aligned}&(m_1 \oplus m_2)(\{x\}) = Km_1(\{x\})m_2(\Theta), (m_1 \oplus m_2)(\{y\})\\ &\quad = Km_1(\{y\})m_2(\Theta),\\ &(m_1 \oplus m_2)(\{z\}) = Km_1(\{z\})m_2(\Theta), (m_1 \oplus m_2)(\{u\})\\ &\quad = Km_1(\Theta)m_2(\{u\}),\\ &(m_1 \oplus m_2)(\{v\}) = Km_1(\Theta)m_2(\{v\}), (m_1 \oplus m_2)(\{w\})\\ &\quad = Km_1(\Theta)m_2(\{w\}),\\ &(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta).\end{aligned}$$

Proof. To combine m_1, m_2, make the intersection table for $m_1 \oplus m_2$ as shown in Table 6. Then, we find that

$$\begin{aligned}1/K = N &= 1 - \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y)\\ &= 1 - m_1(\{x\})m_2(\{u\}) - m_1(\{y\})m_2(\{u\})\\ &\quad - m_1(\{z\})m_2(\{u\} - m_1(\{x\})m_2(\{v\})\\ &\quad - m_1(\{y\})m_2(\{v\}) - m_1(\{z\})m_2(\{v\}\\ &\quad - m_1(\{x\})m_2(\{w\}) - m_1(\{y\})m_2(\{w\})\\ &\quad - m_1(\{z\})m_2(\{w\},\end{aligned}$$

and that m_1, m_2 are combinable if and only if $N > 0$, i.e.,

$$\begin{aligned}&1 - m_1(\{x\})m_2(\{u\}) - m_1(\{y\})m_2(\{u\}) - m_1(\{z\})m_2(\{u\}\\ &\quad - m_1(\{x\})m_2(\{v\}) - m_1(\{y\})m_2(\{v\}) - m_1(\{z\})m_2(\{v\}\\ &\quad - m_1(\{x\})m_2(\{w\}) - m_1(\{y\})m_2(\{w\}) - m_1(\{z\})m_2(\{w\}\\ &\quad > 0,\end{aligned}$$

i.e.,

$$\begin{aligned}&m_1(\{x\})m_2(\{u\}) + m_1(\{y\})m_2(\{u\}) + m_1(\{z\})m_2(\{u\})\\ &\quad + m_1(\{x\})m_2(\{v\}) + m_1(\{y\})m_2(\{v\}) + m_1(\{z\})m_2(\{v\})\\ &\quad + m_1(\{x\})m_2(\{w\}) + m_1(\{y\})m_2(\{w\}) + m_1(\{z\})m_2(\{w\})\\ &\quad < 1.\end{aligned}$$

We know that

$$(m_1 \oplus m_2)(A) = K \sum_{X \cap Y = A} m_1(X)m_2(Y).$$

TABLE 6
Totally Different Points

$m_1 \oplus m_2$	$\{u\}_{m_2(\{u\})}$	$\{v\}_{m_2(\{v\})}$	$\{w\}_{m_2(\{w\})}$	$\Theta_{m_2(\Theta)}$
$\{x\}_{m_1(\{x\})}$	$\{\}_{m_1(\{x\})m_2(\{u\})}$	$\{\}_{m_1(\{x\})m_2(\{v\})}$	$\{\}_{m_1(\{x\})m_2(\{w\})}$	$\{x\}_{m_1(\{x\})m_2(\Theta)}$
$\{y\}_{m_1(\{y\})}$	$\{\}_{m_1(\{y\})m_2(\{u\})}$	$\{\}_{m_1(\{y\})m_2(\{v\})}$	$\{\}_{m_1(\{y\})m_2(\{w\})}$	$\{y\}_{m_1(\{y\})m_2(\Theta)}$
$\{z\}_{m_1(\{z\})}$	$\{\}_{m_1(\{z\})m_2(\{u\})}$	$\{\}_{m_1(\{z\})m_2(\{v\})}$	$\{\}_{m_1(\{z\})m_2(\{w\})}$	$\{z\}_{m_1(\{z\})m_2(\Theta)}$
$\Theta_{m_1(\Theta)}$	$\{u\}_{m_1(\Theta)m_2(\{u\})}$	$\{v\}_{m_1(\Theta)m_2(\{v\})}$	$\{w\}_{m_1(\Theta)m_2(\{w\})}$	$\Theta_{m_1(\Theta)m_2(\Theta)}$

By the intersection table for $m_1 \oplus m_2$, we find that

$$\begin{aligned}
&(m_1 \oplus m_2)(\{x\}) = Km_1(\{x\})m_2(\Theta), (m_1 \oplus m_2)(\{y\}) \\
&\quad = Km_1(\{y\})m_2(\Theta), \\
&(m_1 \oplus m_2)(\{z\}) = Km_1(\{z\})m_2(\Theta), \\
&(m_1 \oplus m_2)(\{u\}) = Km_1(\Theta)m_2(\{u\}), \\
&(m_1 \oplus m_2)(\{v\}) = Km_1(\Theta)m_2(\{v\}), \\
&(m_1 \oplus m_2)(\{w\}) = Km_1(\Theta)m_2(\{w\}), \\
&(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta).
\end{aligned}$$

□

However, now $m_1 \oplus m_2$ is not a 3-points focused mass function, there are focal points $\{x\}, \{y\}, \{z\}, \{u\}, \{v\}, \{w\}$ and Θ. The focusing operator σ should again be applied.

Theorem 5.4.2. *Let m_1, m_2 be two 3-points focused mass functions,*

$$\begin{aligned}
&m_1(\{x\}) + m_1(\{y\}) + m_1(\{z\}) + m_1(\Theta) = 1; \\
&0 \le m_1(\{x\}), m_1(\{y\}), m_1(\{z\}), m_1(\Theta) \le 1; \\
&m_2(\{u\}) + m_2(\{v\}) + m_2(\{w\}) + m_2(\Theta) = 1; \\
&0 \le m_2(\{u\}), m_2(\{v\}), m_2(\{w\}), m_2(\Theta) \le 1.
\end{aligned}$$

Suppose that

$$\begin{aligned}
&m_1(\{x\})m_2(\{u\}) + m_1(\{y\})m_2(\{u\}) + m_1(\{z\})m_2(\{u\}) \\
&\quad + m_1(\{x\})m_2(\{v\}) + m_1(\{y\})m_2(\{v\}) + m_1(\{z\})m_2(\{v\}) \\
&\quad + m_1(\{x\})m_2(\{w\}) + m_1(\{y\})m_2(\{w\}) + m_1(\{z\})m_2(\{w\}) \\
&\quad < 1.
\end{aligned}$$

So, m_1, m_2 are combinable. Denote

$$\begin{aligned}
&(m_1 \oplus m_2)(\{x\}) = Km_1(\{x\})m_2(\Theta) = f(x), \\
&(m_1 \oplus m_2)(\{y\}) = Km_1(\{y\})m_2(\Theta) = f(y), \\
&(m_1 \oplus m_2)(\{z\}) = Km_1(\{z\})m_2(\Theta) = f(z), \\
&(m_1 \oplus m_2)(\{u\}) = Km_1(\Theta)m_2(\{u\}) = f(u), \\
&(m_1 \oplus m_2)(\{v\}) = Km_1(\Theta)m_2(\{v\}) = f(v), \\
&(m_1 \oplus m_2)(\{w\}) = Km_1(\Theta)m_2(\{w\}) = f(w), \\
&(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta),
\end{aligned}$$

where

$$\begin{aligned}
1/K = N = 1 &- m_1(\{x\})m_2(\{u\}) - m_1(\{y\})m_2(\{u\}) \\
&- m_1(\{z\})m_2(\{u\} - m_1(\{x\})m_2(\{v\}) \\
&- m_1(\{y\})m_2(\{v\}) - m_1(\{z\})m_2(\{v\} \\
&- m_1(\{x\})m_2(\{w\}) - m_1(\{y\})m_2(\{w\}) \\
&- m_1(\{z\})m_2(\{w\}.
\end{aligned}$$

Then, $(m_1 \oplus m_2)^\sigma$ is the following:

$$\begin{aligned}
&(m_1 \oplus m_2)^\sigma(\{x'\}) + (m_1 \oplus m_2)^\sigma(\{y'\}) + (m_1 \oplus m_2)^\sigma(\{z'\}) \\
&\quad + (m_1 \oplus m_2)^\sigma(\Theta) = 1, \\
&(m_1 \oplus m_2)^\sigma(\{x'\}) = f(x'),
\end{aligned}$$

where $\{x', y', z', u', v', w'\} = \{x, y, z, u, v, w\}$ and

$$f(x') \ge f(y') \ge f(z') \ge f(u') \ge f(v') \ge f(w'),$$

and $(m_1 \oplus m_2)^\sigma(\{y'\}) = f(y')$, $(m_1 \oplus m_2)^\sigma(\{z'\}) = f(z')$, $(m_1 \oplus m_2)^\sigma(\Theta) = 1 - f(x') - f(y') - f(z')$.

Proof. By Theorem 5.4.1,

$$\begin{aligned}
&(m_1 \oplus m_2)(\{x\}) = Km_1(\{x\})m_2(\Theta) = f(x), \\
&(m_1 \oplus m_2)(\{y\}) = Km_1(\{y\})m_2(\Theta) = f(y), \\
&(m_1 \oplus m_2)(\{z\}) = Km_1(\{z\})m_2(\Theta) = f(z), \\
&(m_1 \oplus m_2)(\{u\}) = Km_1(\Theta)m_2(\{u\}) = f(u), \\
&(m_1 \oplus m_2)(\{v\}) = Km_1(\Theta)m_2(\{v\}) = f(v), \\
&(m_1 \oplus m_2)(\{w\}) = Km_1(\Theta)m_2(\{w\}) = f(w), \\
&(m_1 \oplus m_2)(\Theta) = Km_1(\Theta)m_2(\Theta),
\end{aligned}$$

where

$$\begin{aligned}
1/K = N = 1 &- m_1(\{x\})m_2(\{u\}) - m_1(\{y\})m_2(\{u\}) \\
&- m_1(\{z\})m_2(\{u\} - m_1(\{x\})m_2(\{v\}) \\
&- m_1(\{y\})m_2(\{v\}) - m_1(\{z\})m_2(\{v\} \\
&- m_1(\{x\})m_2(\{w\}) - m_1(\{y\})m_2(\{w\}) \\
&- m_1(\{z\})m_2(\{w\}.
\end{aligned}$$

Then, by the definition of the 3-points focusing operator σ, we find that $(m_1 \oplus m_2)^\sigma$ as follows:

$$\begin{aligned}
&(m_1 \oplus m_2)^\sigma(\{x'\}) + (m_1 \oplus m_2)^\sigma(\{y'\}) + (m_1 \oplus m_2)^\sigma(\{z'\}) \\
&\quad + (m_1 \oplus m_2)^\sigma(\Theta) = 1, \\
&(m_1 \oplus m_2)^\sigma(\{x'\}) = f(x'),
\end{aligned}$$

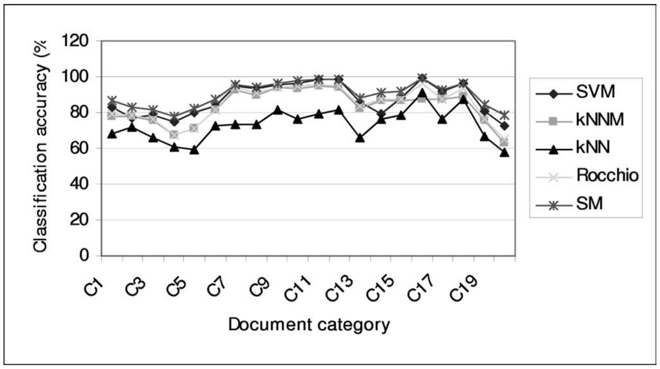


Fig. 3. The performance of the best combined classifier SM (SVM + kNNM) against the individual classifiers SVM, kNNM, kNN, and Rocchio.

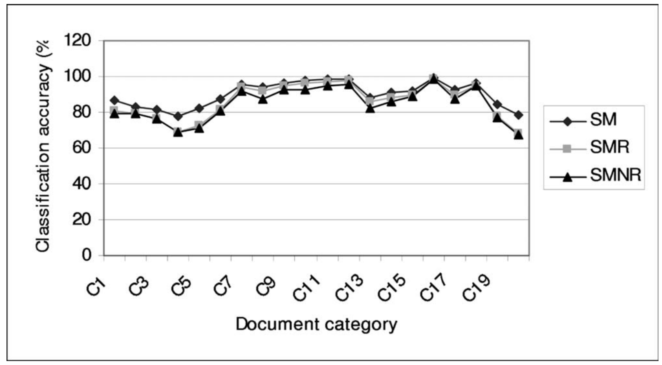


Fig. 4. The performance of the best combined classifier SM (SVM + kNNM) against the individual classifiers SVM, kNNM, kNN, and Rocchio.

where $\{x', y', z', u', v', w'\} = \{x, y, z, u, v, w\}$ and

$$f(x') \geq f(y') \geq f(z') \geq f(u') \geq f(v') \geq f(w'),$$

and $(m_1 \oplus m_2)^{\sigma}(\{y'\}) = f(y')$, $(m_1 \oplus m_2)^{\sigma}(\{z'\}) = f(z')$, $(m_1 \oplus m_2)^{\sigma}(\Theta) = 1 - f(x') - f(y') - f(z')$. □

Notice that we have used Dempster's Rule for the combination of results from two classifiers. A number of properties, including *prima face* weaknesses, of Dempster's rule have been identified, exhaustively analyzed, and dealt with in the literature over the years. We mention a few of these here.

An attractive feature is that, for belief functions, the orthogonal sum gives a result which is independent of the order in which the combinations take place (commutative and associative). Also, a combination of belief functions gives another belief function.

On the other hand, the belief functions to be combined must be based on distinct pieces of evidence. There are strict rules under which the Orthogonal Sum can be used. For example, in the case of TC, we could argue that the pieces of evidence cannot be assumed to be entirely independent and multiple agents methods should be used instead. But, these issues are beyond the scope of the present study. The empirical results of using the orthogonal sum for this purpose have been illustrated in [6].

6 Evaluation

In this section, we describe the experiment which has been performed to evaluate our combination method given in the previous sections. For our experiments, we have chosen a public benchmark data set, often referred to as 20-newsgroup. It consists of 20 categories and each category has 1,000 documents (Usenet articles), so the data set contains 20,000 documents in total. Except for a small fraction of the articles (4 percent), each article belongs to exactly one category [11]. We use *information gain* as a measure for feature selection at the preprocessing stage for each classification method and weight features by using *tfidf* (term frequency within the document and inverse document frequency) after removing function words and applying stemming [19]. In total, 5,300 features have been selected. The experiments have been conducted using a 10-fold cross validation. For each classification method, 10 classifiers are generated and the performance of the method is the mean value of the 10 classifiers. The performance of learning algorithms has been measured using a measure which is widely used in information retrieval and text categorization: the macro-average F_1 defined on a pair of measures, called Precision and Recall [20], [21]. Fig. 3 demonstrates the performance comparison among the best combined classifier (SVM and kNNM—called SM) and four individual classifiers (SVM, kNNM, kNN, and Rocchio) on 20 document categories. The best combined classifier outperforms any individual classifiers on the average. The estimated performance of the best combination is 90.15 percent, which is 2.69 percent better than the best individual classifier (SVM). Fig. 4 illustrates the performance comparison among the best combinations of two classifiers SM (SVM + kNNM), three classifiers SMR (SVM + kNNM + Rocchio), and the four classifiers SMNR (SVM + kNNM + kNN + Rocchio). As we see, the best combination of two classifiers SM outperforms SMR and SMNR and the performance of the best combination of SMR is almost the same as that of SMNR with the exception of document categories of 8-11 and 13-16. The estimated classification accuracies of SMR and SMNR are 86.12 percent and 84.58 percent, respectively, which are 1.35 percent and 2.88 percent worse than the best individual classifier SVM. So, our experimental results show that the combination of the best and the second best classifiers is the best combination that outperforms the individual classifiers and the combined classifiers.

7 Conclusion

In this paper, we have suggested how a novel method and technique for representing outputs from different classifiers—a focal element triplet—can be extended to a focal element quartet. An evidential method for combining multiple classifiers based on this new structure has been analyzed. Similar formulae to those obtained in Section 5 for triplets and quartets can also be obtained for comparison of other numbers of focal elements. The structure and the associated methods and techniques developed in this research are particularly useful for data analysis and decision making under uncertainty.

REFERENCES

[1] Y. Bi, D. Bell, H. Wang, G. Guo, and K. Greer, "Combining Multiple Classifiers Using Dempster's Rule of Combination for Text Categorization," *Proc. First Conf. Modeling Decisions for Artificial Intelligence,* pp. 127-138, 2004.

[2] F. Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Computing Surveys,* vol. 34, no. 1, 2002.

[3] L.S. Larkey and W.B. Croft, "Combining Classifiers in Text Categorization," *Proc. SIGIR-96, 19th ACM Int'l Conf. Research and Development in Information Retrieval,* pp. 289-297, 1996.

[4] Y.H. Li and A.K. Jain, "Classification of Text Documents," *The Computer J.,* vol 41, no. 8, 537-546, 1998.

[5] Y. Yang, T. Ault, and T. Pierce, "Combining Multiple Learning Strategies for Effective Cross Validation," *Proc. 17th Int'l Conf. Machine Learning (ICML '00),* pp. 1167-1182, 2000.

[6] Y. Bi, D. Bell, H. Wang, G. Guo, and K. Greer, "Combining Classification Decisions for Text Categorization: An Experimental Study," *Proc. 15th Int'l Conf. Database and Expert Systems Applications (DEXA '04),* pp. 222-231, 2004.

[7] Y. Bi, D. Bell, and J.W. Guan, "Combining Evidence from Classifiers in Text Categorization," *Proc. Eighth Conf. Knowledge-Based Intelligent Information & Eng. Systems,* pp. 521-528, 2004.

[8] Y. Bi, "Combining Multiple Classifiers for Text Categorization Using Dempster's Rule of Combination," PhD thesis, Univ. of Ulster, 2004.

[9] D.J. Ittner, D.D. Lewis, and D.D. Ahn, "Text Categorization of Low Quality Images," *Proc. Symp. Document Analysis and Information Retrieval,* pp. 301-315, 1995.

[10] Y. Yang, "A Study on Thresholding Strategies for Text Categorization," *Proc. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '01),* pp. 137-145, 2001.

[11] T. Joachims, "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization," *Proc. 14th Int'l Conf. Machine Learning (ICML '97),* 1997.

[12] C.C. Chang and C.J. Lin, "LIBSVM: A Library for Support Vector Machines," http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.

[13] G. Guo, H. Wang, D. Bell, Y. Bi, and K. ieran Greer, "kNN Model-Based Approach in Classification," *Proc. Cooperative Information Systems (CoopIS) Int'l Conf.,* pp. 986-996, 2003.

[14] J.W. Guan and D.A. Bell, " Evidence Theory and Its Applications," *Studies in Computer Science and Artificial Intelligence 7-8,* vols. 1-2, Elsevier, North-Holland, 1991-1992.

[15] T. Mitchell, *Machine Learning.* McGraw-Hill, 1997.

[16] G. Shafer, *A Mathematical Theory of Evidence.* Princeton, N.J.: Princeton Univ. Press, 1976.

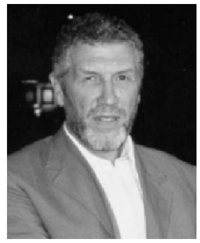
[17] J.-H. Kang and D. Doermann, "Evaluation of the Information-Theoretic Construction of Multiple Classifier Systems," *Proc. Seventh Int'l Conf. Document Analysis and Pattern Recognition,* pp. 789-793, 2003.

[18] L. Kimcheve, J. Bezdek, and R. Duin, "Decision Templates for Multiple Classifier Fusion: An Experimental Comparison," *Pattern Recognition,* vol. 34, pp. 299-314, 2001.

[19] G. Salton, J. Allan, C. Buckley, and A. Singhal, "Automatic Analysis, Theme Generation, and Summarization of Machine-Readable Texts," *Science,* vol. 264, pp. 1421-1426, 1994.

[20] C.J. van Rijsbergen, *Information Retrieval,* second ed. Butterworths, 1979.

[21] S. Zhang, C. Zhang, and Q. Yang, "Information Enhancement for Data Mining," *IEEE Intelligent Systems,* pp. 12-13, Mar./Apr. 2004.



David A. Bell graduated in 1969 in pure mathematics and has three research degrees in a variety of topics: programming language design, database performance, and AI in database systems. He has been a full professor since 1986. He has more than 300 publications, including coauthoring of *Distributed Databases* (Addison-Wesley) and *Evidence Theory and Its Applications* (North Holland), and has supervised approximately 30 PhDs to completion. He has been a prime investigator on many national and EU-funded projects (e.g., MAP, ESPRIT, DELTA, COST, AIM) in IT since 1981. He is a member of the editorial boards of *Information Systems* and *The Computer Journal*, and he has been guest editor of special issues of *Information and Software Technology* and *Information Systems*, on data mining and semantic Web, respectively. He has also chaired /cochaired several program committees, including those for VLDB and ICDE. His research is in data and knowledge management: the linking of reasoning under uncertainty, machine learning, and other artificial intelligence techniques with more established distributed database work exploiting the close relationship between evidence and data.

J.W. Guan has been a visiting professor in the School of Computer Science, The Queen's University of Belfast, United Kingdom, since 2001. He is also a professor in the College of Computer Science and Technology, Jilin University, China. From 1989 to 2000, he was a visiting professor in the School of Information and Software Engineering, University of Ulster, United Kingdom. From 1985 to 1988, he was a visiting professor in the Computer and Information Science Department, University of Massachusetts. From 1980 to 1985, he was a professor in the Computer Science Department at Jilin University, China. He has published about 100 papers and eight books: *Discrete Mathematics*, *The Art of Computer Programming, Linear Automata, Data Structure, Artificial Intelligence, & LISP Programming*, *Principles of Knowledge Engineering, Evidence Theory, and Its Applications, Volume 1-2*. His research interests include computer and information science, reasoning under uncertainty, artificial intelligence, knowledge engineering, and data mining.

Yaxin Bi received the BSc degree in computer science from Jilin University of China, the MSc degree in software engineering from the Graduate School of the Chinese Academy of Science and State Seismological Bureau of China, and the PhD degree in artificial intelligence from the University of Ulster. He is a senior research professional at the Intelligent Systems Research Centre, British Telecom. His main research interests include text mining, machine learning, reasoning under uncertainty, and data and knowledge engineering.