Combining Multiple Hypotheses for Identifying Human Activities

Young-Woo Seo and Katia Sycara

CMU-RI-TR-06-31

May 2006

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

Abstract

Dempster-Shafer theory is one of the predominant methods for combining evidence from different sensors. However, it has been observed that Dempster's rule of combination may yield inaccurate results in some situations. In this paper, we examine the properties and the performance of five different combination rules on a set of real world data. The data was obtained through biometric sensors from a number of human subjects. The problem we study is the prediction of the activity state of a human, given time series readings from the biometric sensors.

Contents

1 Introduction

Information fusion refers to techniques of combining information obtained by multiple sources (homogeneous or heterogeneous) in order to provide comprehensive and accurate information. Merging evidences from different sources would decrease uncertainty in measurement, increase confidence, robustness and coverage. Various techniques, such as Kalman filter [8], [12] and the Dempster-Shafer theory [3], [13] are currently used in the literature for combining evidence. Since the Dempster-Shafer theory is a popular technique for fusing multiple evidences, we want to examine closer the limitations of some of its computational underpinnings.

The Dempster-Shafer theory is a mathematical theory of evidence in which the Shafer's theory of belief functions allows one to quantify partial or incomplete beliefs when it is difficult to estimate complete probability distributions. In the Dempster Shafer theory, a frame of discernment defines the set of all possible outcomes. Different information sources may assign different probabilities on subsets of the same frame of discernment. The Dempster's rule of combination provides a way of combining these multiple assessments through the orthogonal sum. However, as Zadeh pointed out [16], Dempster's rule of combination may yield counterintuitive results because of a normalization factor. In particular, the combination rule emphasizes the agreement between multiple sources and ignores all the conflicts through the normalization factor.

A number of methods and combination rules have been developed to address the problems of the Dempster-Shafer rule of combination in the case of strongly conflicting evidence. In this paper, we examine properties of these different rules of combination and test their performance on a set of real world data. The data is obtained by different biometric sensors that are attached to the human body through an armband and continuously monitor different biological functions, such as temperature, sweat etc. and get readings over time. The problem under study is the prediction of the activity state of a human through fusion of the different sensor readings.

This paper is organized as follows. Section 2 gives a quick overview of the Dempster-Shafer method and presents five different rules of combination. Section 3 describes the experimental setting, data set, and results. Section 4 reviews related work. Section 5 presents conclusions and future work.

2 Overview of Dempster-Shafer Theory

The Dempster-Shafer theory is a mathematical theory of evidence in which Shafer's theory of belief functions allows one to quantify partial or incomplete beliefs. The Dempster's rule of combination provides a way of combining multiple independent hypotheses by the orthogonal sum [9]. It has been widely used to deal with uncertainty, particularly subjective uncertainty that results from the lack of knowledge about a system [15].

The Shafer's model defines a frame of discernment, Θ, as the space of all possible outcomes. A closed-world is assumed in that a frame of discernment is comprised of finite number of exhaustive and exclusive elementary hypotheses. Let Θ be a frame of discernment. For our case, it is a set of human activities that are possibly perceived by deployed sensors. Particularly, let us suppose those human activities are "sleep", "watch TV", and "ambiguous," $\Theta = \{\theta_1, \theta_2, \theta_3\}$, where θ_1 = sleep, θ_2 = watch-tv, θ_3 = ambiguous. The set of all subsets of Θ is defined by the power set of $\Theta, 2^{\Theta}$. Each hypothesis (or subset) will be assigned a subjective belief. In Shafer's model, a basic probability assignment function (bpa, or mass function, m) is used to assign beliefs to subsets of interest. A mass function is a mapping from subsets of a frame of discernment Θ into the unit interval, $m_{\Theta} : 2^{\Theta} \rightarrow [0, 1]$, where, $m_{\Theta}(\phi) = 0$ and $\sum_{A_i \subset \Theta} m_{\Theta}(A_i) = 1$. In contrast to a probability distribution that distributes belief over all possible elements of the outcome space, this distribution attributes belief to subsets of the outcome space. Any subset of Θ that is assigned non-zero mass is called a focal element. The basic probability assignment must lie within an interval, $m(A) \in [bel(A), pl(A)]$. This interval contains the precise probability of a set of interest in the classical probability sense and is bounded by two nonadditive continuous measures called Belief, $bel(A)$, and Plausibility, $pl(A)$. The lower bound (called the degree of belief, $bel(A)$) is the measure of total belief attributed to a hypothesis (or a subset belonging to the frame of discernment), A, where $A \subseteq \Theta$, and accordingly can be calculated by summing all masses (or basic probability assignments) of the proper subsets B that support $A, bel(A) = \sum_{\phi \neq B \subseteq A} m(B)$. The upper bound (called the degree of plausibility, $pl(A)$) quantifies the total amount of belief that might support $A, pl(A) = bel(\Theta) - bel(\bar{A}) = \sum_{B \cap A \neq \phi} m(B)$. Note these two bounds are non-additive, meaning that the sum of all the bel is not required to be 1.

Multiple sources may provide different probabilistic assessments for the same frame of discernment. Dempster's rule of combination aggregates those assessments obtained from multiple independent and equally reliable sources. According to Shafer's model, Dempster's rule of combination fuses multiple beliefs through their basic belief assignments. These beliefs are defined on the same frame of discernment, but are based on independent bodies of evidence.

Dempster rule of combination is performed by using the conjunctive combination rule (also called orthogonal sum).

$$\begin{aligned} m_{12}(A) &= \frac{\sum_{X \cap Y = A} m_1(X) m_2(Y)}{1 - k_{12}} \\ &\text{where,} \\ k_{12} &= \sum_{X \cap Y = \phi} m_1(X) m_2(Y) \end{aligned}$$

where $1 - k_{12}$ is a normalization factor that completely ignores conflict and attributes any belief mass associated with conflict to the null set. Consequently this operation will yield counterintuitive results in the face of significant conflict in certain contexts [16]. In particular, when k_{12} is equal to 1, Dempster's rule of combination cannot be defined mathematically (i.e., divide zero by zero). Thus the combined beliefs do not exist and the bodies of evidences are said to be in full contradiction. We illustrate this through the following example.

Suppose two experts are consulted regarding a system failure. The failure could be caused by either component A, B, or C [16]. The first expert believes that the failure is due to component A with a probability of 0.99 or component B with a probability of

0.01 (denoted by $m_1(A)$ and $m_2(B)$, respectively).

- Expert 1: $m_1(A) = 0.99$ (failure due to component A) and $m_2(B) = 0.01$ (failure due to component B)
- Expert 2: $m_2(B) = 0.01$ (failure due to component B) and $m_2(C) = 0.99$ (failure due to component C).

They agree in their low expectation of failure due to component B, but disagree in the most likely cause.

Table 1: A belief mass matrix is shown.

		m_2		
		$A = 0.99$	$B = 0.01$	C=0
	A = 0	0	0	0
m_1	B=0.01	0.0099	0.0001	0
	C=0.99	0.9801	0.0099	0

$$
\begin{aligned}
m_{12}(B) &= \frac{\sum_{X \cap Y = B} m_1(X) \times m_2(Y)}{1 - \sum_{X \cap Y = \phi} m_1(X) m_2(Y)} \\
&= \frac{0.01 \times 0.01}{1 - ((0.99 \times 0.01) + (0.99 \times 0.99) + (0.01 \times 0.99))} \\
&= \frac{0.0001}{1 - 0.9999} = 1
\end{aligned}
$$

However, if we apply the Dempster's rule of combination to the belief mass matrix in table 1, the basic probability assignment for the failure of component B is 1, even thought there is highly conflicting evidence.

A number of different rules of combination has been developed to address this problem posed by strongly conflicting evidence. In particular, Yager extends Dempster's rule of combination by utilizing the normalization factor as a discounting factor added to the weight of ignorance [14]. Dubois and Prade modifies Dempster's rule of combination based on the assumption that the two sources are reliable when they are not in conflict, but one of them is correct when a conflict occurs [2]. With a slight modification on Shafer's model, Smets [11] and Murphy [6] proposed new rules of combination by simply removing the normalization factor from Dempster's rule of combination. We will detail how those extended rules of combination work in comparison with Dempster's rule of combination in a task of human activity recognition.

2.1 Rules of Combining Multiple Hypotheses

Lefevre and his colleagues proposed a unified framework that can represent several rules of combining multiple hypotheses in a conjunctive consensus [5]. Their idea generalizes previously proposed rules of combination based on Shafer's model in the following two steps: first, compute the total conflicts in terms of the conjunctive consensus:

$$k_{12} = \sum_{X \cap Y = \phi} m_1(X) m_2(Y)$$

where $X, Y \in 2^{\Theta}$. And then, second, combine hypotheses on the subset of the frame of discernment, $(A \neq \phi) \subseteq \Theta$ with a particular set of coefficients, $w_m(A) \in [0, 1]$:

$$\begin{aligned} m(\phi) &= w_m(\phi) k_{12} \\ m(A) &= \sum_{X \cap Y = A} m_1(X) m_2(Y) + w_m(A) k_{12} \end{aligned}$$

where $\forall (A \neq \phi) \in 2^{\Theta}$ and $\sum_{A \subseteq \Theta} w_m(A) = 1$. By using this unified framework, one can represent a particular rule of combination by choosing a particular set of coefficients.

- Yager recognizes that an unreliable conclusion results from the conflicts between hypotheses and extends Dempster's rule of combination by utilizing k_{12} as a discounting factor added to the weight of ignorance [14]. Yager's rule of combination is obtained by choosing $w_m(\Theta) = 1$ and $w_m(A \neq \Theta) = 0$

$$\begin{aligned} m(\phi) &= 0 \\ m(A) &= \sum_{X \cap Y = A} m_1(X) m_2(Y), \\ m(\Theta) &= m_1(\Theta) m_2(\Theta) + \sum_{X \cap Y = \phi} m_1(X) m_2(Y) \\ & \quad \text{if } A = \Theta \end{aligned}$$

  where $\forall A \in 2^{\Theta}, A \neq \phi, A \neq \Theta$.

- Murphy's rule of combination involves a simple arithmetic average of belief functions associated with two independent evidences [6].

$$m(A) = \frac{1}{2} \left[m_1(A) + m_2(A) \right]$$

- Dubois and Prade assume that the two sources are reliable when they are not in conflict, but one of them is correct when a conflict occurs [2]. In other words, the truth lies in $X \cap Y$ if $X \cap Y \neq \phi$ or $X \cup Y$ if $X \cap Y = \phi$. Their rule of combination is a hybrid approach that is a trade-off between precision and reliability among the multiple sources. This rule is obtained by choosing $\forall A \subseteq P$ and $w_m(A) = \frac{1}{1-k_{12}} \sum_{A_1, A_2 | A_1 \cup A_2 = A, A_1 \cap A_2 = \phi} m^*$

$$\begin{aligned} m(\phi) &= 0 \\ m(A) &= \sum_{X \cap Y = A, X \cap Y \neq \phi} m_1(X) m_2(Y) + \\ & \quad \sum_{X \cup Y = A, X \cap Y = \phi} m_1(X) m_2(Y) \end{aligned}$$

where $\forall A \in 2^{\Theta}, A \neq \phi$.

- Smets modifies Dempster's rule of combination by removing the normalization factor, k_{12} [11]. His rule also allows positive mass on the null/empty set ϕ. Smets' rule of combination is obtained by choosing $w_m(\phi) = 1$ and $w_m(A \neq \phi) = 0$:

$$\begin{aligned} m(\phi) &= k_{12} = \sum_{X \cap Y = \phi} m_1(X) m_2(Y) \\ m(A) &= \sum_{X \cap Y = A} m_1(X) m_2(Y) \end{aligned}$$

where $\forall (A \neq \phi) \in 2^{\Theta}$.

3 Experiments

Our task is to develop a reliable method that identifies a human activity from given sensor readings. A single electronic device (i.e., armband) including multiple sensors is attached to human body, measuring various body movement, e.g., total energy expenditure, physical activity duration, type of physical activity. The device has a model of human activity for recognizing a possible activity from a sensory report, which is generated by multiple sensors at a discrete time step. Given collected sensor data, one might want to train multiple classifiers and to generate multiple models, hoping that the performance of trained classifiers is boosted by synergistically integrating their outputs. To this end, we want to test the effectiveness of several different rules of combination.

3.1 Human Body Sensor Data

We used a collection of human body sensor data that was collected from sensory armband attached to several human subjects. This data set was originally released to a Physiological Data Modeling Contest (PDMC) [1]. An armband attached to human body monitors a subject's motion, skin response, and heart's flux/beats thru 9 different sensors over time. In particular, each sensor reading is comprised of 16 attributes

- Demographic information: age, handedness, gender,
- Sensory information:
  - $Sensor_1$ measures sweat in micro siemens,
  - $Sensor_2$ measures heat lost to the environment in watts/meter^2,
  - $Sensor_3$ measures temperature near armband in degrees centigrade,
  - $Sensor_4$ measures number of steps,
  - $Sensor_5$ measures temperature of skin in contact with armband in degrees centigrade,

[1] It is publicly available at http://www.cs.utexas.edu/users/sherstov/pdmc/

- $Sensor_6$ measures sum of absolute differences of vertical acceleration in gram,
- $Sensor_7$ measures average of vertical acceleration in gram,
- $Sensor_8$ measures sum of absolute differences of horizontal acceleration in gram,
- $Sensor_9$ measures average of horizontal acceleration,

• User information: user ID, sessional ID, session time

There are four different contexts about human activities: "sleep", "sleep-unlabeled activities", "watch-tv", and "watch-tv-unlabeled activities." In particular, an activity annotated as "sleep-unlabeled" means that a human subject is sleeping while doing other unknown (or unlabeled) activities.[2]

3.2 Experimental Setting

In our scenario, there are k different experts that identify a possible human activity given a sensor reading s_j. Each expert corresponds to an electronic device attached to human body. An expert has the frame of discernment and a mass function of sensor reading. In particular, the mass function is a mapping from subsets of the frame into the unit interval, $h_i(s_j) : 2^\Theta \rightarrow [0, 1]$, where $m_\Theta(\phi) = 0$ and $\sum_{A_i \subset \Theta} m_\Theta(A_i) = 1$. The frame of discernment includes three exclusive and exhaustive elements about possible human activities, $\Theta = \{\theta_1, \theta_2, \theta_3\}$, where θ_1 = *sleep*, θ_2 = *watch-TV*, θ_3 = *unlabeled* activity. An unlabeled activity is one where human subjects do not specify their activity while they took part in the data collection. When an expert receives a sensory reading, it assigns its belief to the following four belief masses, $m(\theta_1), m(\theta_2), m(\theta_1 \cup \theta_3)$, and $m(\theta_2 \cup \theta_3)$. Since we are interested in identifying only those four activities, the corresponding subsets are chosen from the power set, $2^\Theta = \{\phi, \theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \theta_1 \cup \theta_2 \cup \theta_3\}$. For example, $m(\theta_1)$ is a belief mass where an expert believes that the human subject is sleeping whereas $m(\theta_2 \cup \theta_3)$ represents the degree of an expert's belief on the hypothesis that a human subject is either watching TV or watching TV while doing other unknown activities.

There are more than 100,000 labeled sensory readings that are collected from 49 human subjects (i.e., 12 men and 7 women in the training set and 21 men and 9 women in the test set). Instead of using all the available sensor data, we randomly chose a subset of them and used for our experiments. A subset is approximately 10% of the original data set and is divided into two parts: 80% for training set and 20% for test set. For a training data of each subset, k different classifiers are obtained by using linear regression.

$$
\begin{aligned}
\mathbf{y} &= \mathbf{X}\beta + \epsilon, (\epsilon = RSS(\beta) = \mathbf{y} - \mathbf{X}\beta) \\
\hat{\mathbf{y}} &= \mathbf{X}\hat{\beta} \\
& \text{where,}
\end{aligned}
$$

[2] The *other* activities are ones without correct annotations. The *sleep* and *watch-tv* are two clearly known when the data set was released.

$$
\begin{aligned}
\hat{\beta} &= \min_{\beta}\{\epsilon^2\} = \min_{\beta}\left\{(\mathbf{y}-\mathbf{X}\beta)^T(\mathbf{y}-\mathbf{X}\beta)\right\} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}
\end{aligned}
$$

where $\mathbf{y}$ is $n \times 1$ vector of the target values, $\mathbf{X}$ is a $n \times |\text{Sensor Reading}|$ sampled subset matrix, β is a $|\text{Sensor Reading}| \times 1$ coefficient vector, and $|\text{Sensor Reading}|$ is 9 (9 sensor readings). $\hat{\beta}$ is the least squares estimate that minimizes the residual sum of squares, $RSS(\beta) = min_{\beta}(y - X\beta)^T(y - X\beta)$. Since we cannot always guarantee that X is nonsingular and X^TX is positive definite, we utilize a singular value decomposition when to estimate the pseudo inverse, $(X^TX)^{-1}X^T$. [3]

For the testing data of each subset, each of k trained classifiers first assigns its beliefs to four possible masses, $m(\theta_1), m(\theta_2), m(\theta_1 \cup \theta_3), m(\theta_2 \cup \theta_3)$, and then returns the mass of the largest value as the most probable hypothesis about a given sensory reading. Accordingly there are k different hypotheses from the corresponding number of experts on a sensor report. We then compare the effectiveness of five different rules of combination – which rule is more effective than others in combining those different hypotheses. The value of k is empirically determined in the next section.

Table 2: Per-activity contingency table. *a* is true positive, *b* is false negative, *c* is false negative, and *d* is true negative, respectively.

		True	
		activity = 1	activity = 0
Output	activity = 1	a	b
	activity = 0	c	d

3.3 Experimental Results

Table 2 shows a contingency table. A contingency table is built for each of four masses and used to quantify the performance of a rule of combination. For instance, suppose we build a contingency table for *sleep* activity. *a* is true positive if a sensor report, s_j, truly belongs to *sleep* activity and is classified as *sleep* activity. *c* is false negative, i.e. whereas the true activity is *sleep*, s_j is classified as other activity (e.g., *watch-tv*). *d* is true negative if s_j truly belongs to other activity (e.g., *watch-tv*) and is classified as that activity (e.g., *watch-tv*). Finally *b* is false positive if s_j is classified as *sleep* whereas in reality it is *watch-tv*. With these quantifiers, we utilize six different measures in order to detail the performance;

- Accuracy, $acc = \frac{a+d}{a+b+c+d}$, if $a + d > 0$, otherwise undefined,
- Precision, $p = \frac{a}{a+b}$, if $a > 0$, otherwise undefined,
- Recall, $r = \frac{a}{a+c}$, if $a > 0$, otherwise undefined,

[3] $\hat{\beta} = V_k\Sigma_k^{-1}U_k^T$ where $X = U\Sigma V^T$ and k is the rank of X.

- F1, $f1 = \frac{2rp}{r+p} = \frac{2a}{2a+b+c}$, if $(a+b+c) > 0$, otherwise undefined,
- Miss (or false negative), $m = \frac{c}{a+c}$, if $c > 0$, otherwise undefined,
- False alarm (or false positive), $f = \frac{b}{b+d}$, if $b > 0$, otherwise undefined.

The macro-average is used to measure global performance because there are four different human activities to recognize. The macro-average is obtained by producing per-activity performance measure first and then averaging the corresponding measures. Table 3 shows the performance of rules of combination in macro-average.

Table 3: Experimental results are presented in macro-average over four different human activities. The values in bold face are the best for corresponding performance measure. Due to limited space, the names of methods and measures are acronymized. *D* is Dempster's rule of combination, *Y* is Yager's, *M* is Murphy's, *DP* is Dubois and Prade's, *S* is Smet's rule of combination. "Base" represents a base line method that simply adds the corresponding beliefs of multiple hypothesis and returns the largest mass. *p* is precision, *r* is recall, *m* is miss, and *f* is false alarm, respectively. The higher the better in the precision, recall, and f1 measures whereas the lower the better in the miss and false alarm.

	Base	D	Y	M	DP	S
acc	0.773	0.779	**0.798**	0.771	0.773	0.773
p	0.661	0.542	0.581	**0.669**	0.659	0.659
r	0.516	0.536	**0.567**	0.514	0.515	0.515
f1	0.449	0.477	**0.564**	0.455	0.447	0.447
m	0.483	0.463	**0.432**	0.485	0.484	0.484
f	0.151	0.144	**0.135**	0.153	0.152	0.152

The accuracy results were best Yager's, Dempster's, base-line, Dubois and Prade's, Smet's, and finally, Murphy's. In general, the performance does not vary distinctively because, for our task, no conflicts among hypotheses is expected to happen. In other words, there were always four focal points that our four masses (i.e., $m(\theta_1), m(\theta_2), m(\theta_1 \cup \theta_3), m(\theta_2 \cup \theta_3)$) are always greater than zero. This caused the modified rules of combination to produce quite similar results to those of Dempster's.

Yager's rule showed the best performance on all measures except for the precision because it uses the normalization factor for a discounting factor added to the weight of ignorance. There is approximately 2.5% improvement achieved by Yager's rule of combination over the base line method. This is quite similar to that of Bi and his colleagues's finding that 3% improvement was obtained by combining heterogeneous non-parametric classifiers [1]. Surprisingly the base line performed relatively well. The precision of the base line method is actually better than that of Yager's. However Dempster's rule of combination did not show a good result because primarily it completely ignores all the conflicts among hypotheses through the normalization factor. It is interesting to observe that Murphy's rule showed competent results, even though it is relatively simple to combine multiple hypotheses.

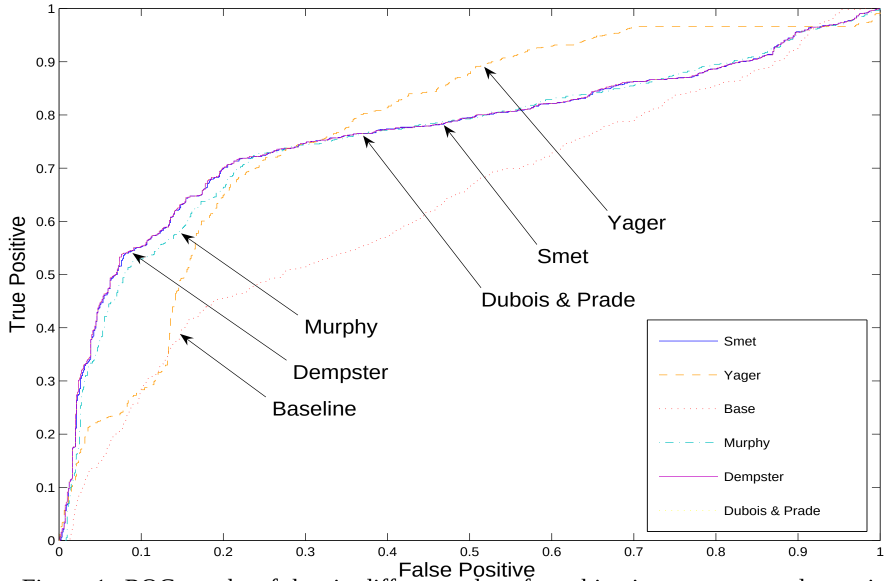

Figure 1: ROC graphs of the six different rules of combination are presented. x-axis represents false positive rates and y-axis represents true positive rates.

Figure 1 shows ROC (Receiver Operating Characteristics) graphs of the six different methods. A ROC curve is used to depict the tradeoff between true positive rates (or hit rates) plotted on y-axis and false positive rates (or false alarm rates) on x-axis. A point in ROC space is better than another if it is to the northwest, i.e. true positive rate is higher, false positive rate is lower. ROC graphs by Dubois/Prade is almost identical to that of Smets' because for our application these two rules did not produce different results. In particular, a combined hypothesis about θ_1 by Dubois and Prade rule is $m_{12}(\theta) = \sum_{X \cap Y = \theta_1} m_1(X)m_2(Y) + \sum_{X \cup Y = \theta_1, X \cap Y = \phi} m_1(X)m_2(Y)$. Since the second term, $\sum_{X \cup Y = \theta_1, X \cap Y = \phi} m_1(X)m_2(Y)$ is zero, the result by Dubois and Prade rule is the same as that of Smet's rule, $m_{12}(\theta) = \sum_{X \cap Y = \theta_1} m_1(X)m_2(Y)$, where $\forall(\theta_1 \neq \phi) \in 2^{\Theta}$. These identical results are also observed in table 3.

4 Related Work

The idea of combining multiple evidences (or hypotheses) is commonly accepted in the fusion community since a decision made on the basis of the multiple pieces of evidence should be more effective than one based on single piece of evidence.

Dempster's rule of combination has been widely used to combine multiple source of evidences. Girondel and his colleagues applied Dempster-Shafer theory to the task of recognizing human body postures [3]. The task is to recognize one of four predefined postures based on data about people 2D segmentation and their face localization. Dempster-Shafer theory is used to enhance limits of different classifiers by combining their results. Wu and his colleagues utilized Dempster-Shafer theory for fusing sensor data in tracking a user's focus of attention [13]. Bi and his colleagues [1] investigate the usefulness of Dempster's rule of combination in aggregating four different text classification methods; support vector machines, non-parametric k-nearest neighbors, parametric k-nearest neighbors, and Rocchio methods. They found that there is approximately 3% improvement achieved by combining multiple classifiers over the best individual method on average. Hu and Damper [4] proposed a heuristic based on the weighted sum rule to improve the performance of speaker identification task. In particular, their method chose weighting parameters for classifiers that maximize the correct identification estimation functions, assuming that there are multiple Gaussian distributions behind the utterance generation processes. Quost and his colleagues reviewed several methods for combining pairwise classifiers in the transferable belief model (TBM) [7]. They found that the TBM is flexible in assessing the validity of the classifiers while fusing the results of multiple classifiers. Smarandache and Dezert reviewed 6 different rules of combination and compared their effectiveness with a couple of numerical examples [10]. However since their examples are used for an illustrative purpose, it is difficult to distinguish their possible difference in real-world applications.

5 Conclusions and Future Work

In this paper, we examined the performance of six rules (including the Dempster-Shafer rule) of combination that have been proposed in the literature to mitigate shortcomings of the Dempster-Shafer rule of combination. We tested these rules on biometric data obtained through sensor readings from different humans. We measured the performance of the different rules according to measures, such as precision, and recall, as well as by examining the ROC curves. Our results show that Yager's rule outperformed all others. The results of this study encourage the more widespread use of Yager's rule.

The Dempster-Shafer theory considers a static frame of discernment that is known a priori. However, it may happen that all possible outcome classes are not known a priori. In future work, we would like to develop a method that can combine evidence dynamically while refining the frame of discernment on the fly.

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. In *Proceedings of Modeling Decisions for Artificial Intelligence*, pages 127–138, 2004.

[2] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. 4:244–264, 1988.

[3] V. Girondel, L. Bonnaud, A. Caplier, and M. Rombaut. Belief theory-based classifiers comparison for static human body postures recognition in video. 2(1):29–33, 2005.

[4] R. Hu and R. Damper. Fusion of two classifiers for speaker identification: Removing and not removing silence. In *Proceedings of International Conference on Information Fusion (Fusion-05)*, 2005.

[5] E. Lefevre, O. Colot, and P. Vannoorenberghe. Belief functions combination and conflict management. 3(2):149–162, 2002.

[6] C. Murphy. Combining belief fuctions when evidence conflicts. 29:1–9, 2000.

[7] B. Quost, T. Denoeux, and M. Masson. Pairwise classifier combination in the transferable belief model. In *Proceedings of International Conference on Infomation Fusion (Fusion-05)*, 2005.

[8] S. Roumeliotis and G. Bekey. An extended kalman filter for frequent local and infrequent global sensor data fusion. In *Proceedings of the Sensor Fusion and Decentralized Control in Autonomous Robotic Systems (SPIE-97)*, pages 11–22, 1997.

[9] G. Shafer. *A Mathematical Theory of Evidence*. Princeton Univ. Press, 1976.

[10] F. Smarandache and J. Dezert. A simple proportional conflict redistribution rule. 3(J05):1–36, 2005.

[11] P. Smets and R. Kennes. The transferable belief model. 66(2):191–234, 1994.

[12] S. Sun and Z. Deng. Multi-sensor information fusion kalman filter weighted by scalars for systems with colored measurement noises. 127(4):663–667, 2005.

[13] H. Wu, M. Siegel, and S. Ablay. Sensor fusion using dempster-shafer theory ii: Static weighting and kalman filter-like dynamic weighting. In *Proceedings of IMTC'2003*, 2003.

[14] R. Yager. Hedging in the combination of evidence. 4(1):73–81, 1983.

[15] R. Yager, J. Kacprzyk, and M. Fedrizzi. *Advances in the Dempster-Shafer Theory of Evidence*. Wiley, 1994.

[16] L. A. Zadeh. Review of books: A mathematical theory of evidence. 5(3):81–83, 1984.