Classification with Imprecise Likelihoods: A comparison of TBM, Random Set and Imprecise Probability Approach

Alessio Benavoli
IDSIA
Manno, Switzerland,
alessio@idsia.ch

Branko Ristic
DSTO
Melbourne, Australia
branko.ristic@dsto.defence.gov.au

***Abstract*— The problem is target classification in the circumstances where the likelihood models are imprecise. The paper highlights the differences between three suitable solutions: the Transferrable Belief model (TBM), the random set approach and the imprecise probability approach. The random set approach produces identical results to those obtained using the TBM classifier, provided that equivalent measurement models are employed. Similar classification results are also obtained using the imprecise probability theory, although the latter is more general and provides more robust framework for reasoning under uncertainty.**

Keywords: Model-based classification, imprecise likelihoods.

I. Introduction

The problem is classification based on imprecise probabilistic models. The recent paper [1] demonstrated that the standard Bayes classifier in these circumstances behaves in a manner that is inconsistent with our intuition. The main difficulty is that the standard Bayes classifier can work only with precise likelihoods. For that reason it needs to forcefully adopt the precise likelihoods, thereby ignoring the uncertainty in mapping the class to the feature measurement.

As a more appropriate alternative to the standard Bayes classifier, [1] proposed a classification method based on the transferrable belief model (TBM) [2]. The main idea of this method is to treat the precise likelihoods which are used by the standard Bayes classifier, as subjective, rather than true, models. The precise likelihoods are therefore converted to the least-committed belief functions, followed by the application of the generalised Bayes transform [3] for classification. The final step involves the mapping from the belief domain into the probabilistic domain via the pignistic transform [2]. The numerical results in [1] confirmed that the TBM classifier indeed behaves in accordance with our intuition.

Mahler [4, Ch.4-8] recently proposed a novel approach to Bayesian estimation, fusion and classification, applicable to situations where the information (priors, measurements, likelihoods) is imprecise and vague (fuzzy) in addition to being random. Mahler refers to this type of information as *non-traditional* data [4], [5]. In his approach, the uncertainty inherent in non-traditional data is represented by a Random Set (RS) (rather than random variables). He argues that the Bayesian framework for inference combined with random set representation of non-traditional data is the appropriate approach to reasoning under uncertainty.

As shown by Walley [6], belief functions [2], [7], random sets [4] and possibility measures [8] represent uncertainty through sets instead of single probability measures, and they can be regarded as special cases of Walley's *coherent lower previsions* [6]. This theory, which is usually referred to as *imprecise probability* theory, provides a very general model of uncertain knowledge. In this approach, uncertainty inherent data (likelihood model) and/or prior information (prior model) is represented by set of probability measures or, equivalently, by coherent lower/upper previsions (expectations). Stemming from de Finetti's work on subjective probability, Walley adopts a behavioural interpretation of probability. His theory is based on three mathematical concepts: avoiding sure loss, coherence and natural extension. A probabilistic model avoids sure loss if it cannot lead to behaviour that is certain to be harmful. Coherence is a stronger principle which characterizes a type of self-consistency. Natural extension allows to build coherent models from any set of probability assessments that avoid sure loss. Walley's theory is directly applicable to both discrete and continuous state domain. For all these reasons, Walley's theory is a solid framework for modelling uncertainty in information fusion problems [9].

The aim of this paper is to highlight the differences between the TBM, RS and imprecise probability approaches for reasoning. A target classification problem with imprecise likelihood functions is adopted as a simple and illustrative example for practical comparison. As it was argued in [10], the RS approach produces identical results to those obtained using the TBM classifier, provided that equivalent measurement models are employed. The paper shows that similar classification results can be obtained using the imprecise probability theory. However, since the imprecise probability theory is more general, it provides more robust results and is able to deal with more general cases than both TBM and RS approach.

II. Imprecise probability

This section briefly reviews Walley's theory [6]. Consider a variable X taking values in a set $\mathcal{X}$ and a utility (bounded) function $f : \mathcal{X} \to \mathbb{R}$, with $f \in \mathcal{L}(\mathcal{X})$, where $\mathcal{L}(\mathcal{X})$ denotes the linear space of all utilities on $\mathcal{X}$ [6, Ch.1]. Let $p : \mathcal{X} \to \mathbb{R}$ be the probability mass function (PMF) or the probability density function (PDF) of X, then the expected utility of f w.r.t. $p(\cdot)$ is defined as:

$$E_X[f] = \int\limits_{x\in\mathcal{X}} f(x)p(x)dx \tag{1}$$

where $p(x)$ denotes the value that the PDF of X assumes for $X = x$. Notice that the integral becomes a sum in the case $p(\cdot)$ is a PMF. Consider now the case where we have incomplete information to elicit a single distribution to describe the probabilistic information over X. Suppose we can characterize this incomplete information by means of a closed-convex set of PDFs $K(X)$. For each PDF $p \in K(X)$, we can compute the expectation $E_X[f]$. Thus, because of linearity of expectation operator, it results that: $\underline{E}_X[f] \leq E_X[f] \leq \overline{E}_X[f]$, where $\underline{E}_X[f]$ is the so called coherent lower prevision (CLP) or lower expectation of f which, by definition, is equal to:

$$\underline{E}_X[f] = \min_{p\in K(X)} \int\limits_{x\in\mathcal{X}} f(x)p(x)\ dx, \quad \forall f \in \mathcal{L}(\mathcal{X}) \tag{2}$$

and $\overline{E}_X[f] = -\underline{E}_X[-f]$ is the upper expectation.

Let us present some examples of CLPs. CLP in (1) corresponds to the most informative one, i.e., the case in which $K(X)$ includes only one density and, thus, $\overline{E}_X[f] = \underline{E}_X[f] = E_X[f]$. A CLP $\underline{E}_X$ on $\mathcal{L}(\mathcal{X})$ such that $\underline{E}_X[f] = \min_{x\in\mathcal{X}} f(x)$ for all $f \in \mathcal{L}(\mathcal{X})$ can be easily identified as the least informative CLP and is therefore called *vacuous*. In fact, in the case $K(X)$ includes all possible densities, (2) reduces to $\underline{E}_X[f] = \min_{x\in\mathcal{X}} f(x)$.[1]

Now consider also a second variable Z with values in $\mathcal{Z}$. For each utility $h \in \mathcal{L}(\mathcal{X}\times\mathcal{Z})$ and $z \in \mathcal{Z}$, a *conditional lower prevision* $\underline{E}_Z(h|X = x)$, denoted also as $\underline{E}_Z(h|x)$, is defined as the lower expectation of h w.r.t. Z conditioned on the case that the variable X assumes the value x. $\underline{E}_Z(h|x)$ can be obtained from definition (2) by replacing $p(x) \in K(X)$ with the conditional PDFs $p(z|x) \in K(Z|X)$, where $K(Z|X)$ is the set of conditional densities.

Given a conditional CLP $\underline{E}_Z[\cdot|X]$ and a marginal CLP $\underline{E}_X$, a joint CLP on $\mathcal{L}(\mathcal{X}\times\mathcal{Z})$ can be obtained by *marginal extension*:

$$\underline{E}_{X,Z}[h] = \underline{E}_X\left[\underline{E}_Z[h|X]\right],\ \forall h \in \mathcal{L}(\mathcal{X}\times\mathcal{Z}). \tag{3}$$

On the other hand, given a CLP $\underline{E}_{X,Z}$ on $\mathcal{L}(\mathcal{X}\times\mathcal{Z})$ we can derive the conditional CLP $\underline{E}_X[\cdot|z]$ as follows:

$$\underline{E}_X[f|z] = \min_{p\in K(X,Z)} \frac{\int\limits_{x\in\mathcal{X}} f(x)p(x,z)dx}{p(z)},\ \forall f \in \mathcal{L}(\mathcal{X}) \tag{4}$$

where it has been assumed that $p(z) = \int_{x\in\mathcal{X}} p(z|x)p(x) > 0$ with $p(x,z) = p(z|x)p(x)$. Notice that $\underline{E}_X[f|z]$ is equal to the lower envelope of all conditional expectations which can be obtained by applying Bayes rule to each density in $K(X,Z)$. For this reason, (4) is called *Generalized Bayes Rule* (GBR) [6].[2] Notice that, in (4), instead of considering all the densities it is sufficient to consider the extreme densities of $K(X,Z)$ [6, Sec. 6.4.2], i.e., the solution of GBR is on the extremes of $K(X,Z)$.

A. Imprecise probability and belief functions

Consider a closed-convex set of probability mass functions and its associated CLP computed over indicators of subsets of $\mathcal{X}' \subseteq \mathcal{X}$. That is, given a set $\mathcal{X}' \subseteq \mathcal{X}$, we consider the lower expected value of the indicator function $I_{\{\mathcal{X}'\}}$, which, by definition, is equal to the lower probability that $X \in \mathcal{X}'$, i.e.:

$$\begin{aligned} \underline{E}_X[I_{\{\mathcal{X}'\}}] &= \min_{p\in K(X)} \sum_{x_j\in\mathcal{X}} I_{\{\mathcal{X}'\}}(x_j)p(x_j) \\ &= \min_{p\in K(X)} \sum_{x_j\in\mathcal{X}'} p(x_j) = \underline{P}_X(\mathcal{X}'), \end{aligned} \tag{5}$$

where $\underline{P}_X(\mathcal{X}')$ is used to denote the lower probability of the event $X \in \mathcal{X}'$. If $\underline{P}_X$ satisfies these properties:

1) $\underline{P}_X(\emptyset) = 0$, $\underline{P}_X(\mathcal{X}) = 1$,
2) for every positive integer n and every collection $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ of subsets of $\mathcal{X}$,

$$\underline{P}_X\left(\bigcup_i \mathcal{X}_i\right) \geq \sum_{\substack{I\subseteq\{1,\ldots,n\},\\ I\neq\emptyset}} (-1)^{|I|-1}\underline{P}_X\left(\bigcap_{i\in I}\mathcal{X}_i\right) \tag{6}$$

where $|I|$ denotes the cardinality of the set I, then the lower probability $\underline{P}_X(\cdot)$ is a belief function and its conjugate upper probability $\overline{P}(\cdot)$ is a plausibility function.[3] Property (6) is commonly referred to as ∞-monotonicity. Only lower probabilities that satisfy this property are belief functions.[4] A belief function is thus a special case of CLPs [6].

Given a belief function, the extreme PMFs of the set $K(X)$ that generate the belief function can be obtained as follows [11]:

$$p_{\pi^\ell}(x_{\pi_i^\ell}) = \sum_{\substack{\mathcal{X}'\subseteq\mathcal{X},\ s.t.:\\ x_{\pi_i^\ell}\in\mathcal{X}',\, x_{\pi_j^\ell}\notin\mathcal{X}',\forall j<i}} m(\mathcal{X}'), \tag{7}$$

where π^ℓ is one of the possible permutations of the elements of $\mathcal{X} = \{x_{\pi_1^\ell}, \ldots, x_{\pi_n^\ell}\}$ and m is the basic belief assignment

[1] This can be proven by noticing that, when $K(X)$ includes all possible densities, the PDF which obtains the minimum $\min\limits_{p\in K(X)} E_X[f]$ is $p(x) = \delta(x-\tilde{x})$, where δ is a Dirac delta in $\tilde{x} = \arg\min_{x\in\mathcal{X}} f(x)$.

[2] In [6, Ch. 6], GBR is defined directly in terms of $\underline{E}_{X,Z}$ without the needs of explicitly determining $K(X,Z)$. Both the definitions are equivalent under standard assumptions of applicability of Bayes rule.

[3] They coincide with the definition of belief and plausibility functions given in Dempster-Shafer theory.

[4] When symbol $\geq$ in (6) is replaced by equality, then ∞-monotonicity reduces to the well-known additivity rule for non-exclusive events in standard probability. Probability functions are in fact a special case of belief functions.

(bba) relative to the belief function. Each probability function in (7) assigns to each singleton $x_{\pi_i^\ell}$ the mass of all the focal elements of the bba which contain it, but do not contain the elements which precede $x_{\pi_i^\ell}$ in the ordered list $\{x_{\pi_1^\ell}, \ldots, x_{\pi_n^\ell}\}$ generated by the permutation π^ℓ.

Consider, for instance, the belief function in Table I. The possible permutations of the elements of the possibility space $\mathcal{X} = \{x_1, x_2, x_3\}$ are:

$$\pi^1 = \{x_1, x_2, x_3\},\ \pi^2 = \{x_1, x_3, x_2\},\ \pi^3 = \{x_2, x_1, x_3\},$$
$$\pi^4 = \{x_2, x_3, x_1\},\ \pi^5 = \{x_3, x_1, x_2\},\ \pi^6 = \{x_3, x_2, x_1\}.$$

Consider then the first permutation π^1, this means that $x_{\pi_1^1} = x_1$, $x_{\pi_2^1} = x_2$, $x_{\pi_3^1} = x_3$. From the bba m in Table I and (7) with $i = 1$ and $\ell = 1$, one gets:

$$p_{\pi^1}(x_{\pi_1^1}) = p_{\pi^1}(x_1) = \sum_{x_1 \in \mathcal{X}'} m(\mathcal{X}') = m(x_1) + m(x_1, x_2) + m(x_1, x_3) + m(x_1, x_2, x_3) = 0.6.$$

In fact in this case $\{x_{\pi_j^1} \notin \mathcal{X}', \forall j < i = 1\}$ is an empty set, since there are not elements before x_1 in the permutation π^1. Conversely, for $i = 2$, it holds that $\{x_{\pi_j^1} \notin \mathcal{X}', \forall j < i = 2\} = \{x_1\}$ and thus:

$$\begin{aligned} p_{\pi^1}(x_{\pi_2^1}) = p_{\pi^1}(x_2) &= \sum_{x_2 \in \mathcal{X}',\ x_1 \notin \mathcal{X}'} m(\mathcal{X}'), \\ &= m(x_2) + m(x_2, x_3) = 0.35, \end{aligned}$$

and finally $p_{\pi^1}(x_3) = m(x_3) = 0.05$. For the permutation π^2, it still results that $p_{\pi^1}(x_{\pi_1^1}) = p_{\pi^2}(x_1) = 0.6$, since x_1 is the first element of π^2, but being $x_{\pi_2^2} = x_3$ one has:

$$\begin{aligned} p_{\pi^2}(x_{\pi_2^2}) = p_{\pi^2}(x_3) &= \sum_{x_3 \in \mathcal{X}',\ x_1 \notin \mathcal{X}'} m(\mathcal{X}'), \\ &= m(x_3) + m(x_2, x_3) = 0.3. \end{aligned}$$

and $p_{\pi^2}(x_{\pi_3^2}) = p_{\pi^2}(x_2) = 0.1$. Consider the permutation π^5, in this case $p_{\pi^5}(x_{\pi_1^5}) = p_{\pi^1}(x_3)$ is equal to:

$$\begin{aligned} p_{\pi^1}(x_3) &= \sum_{x_3 \in \mathcal{X}'} m(\mathcal{X}') = m(x_3) + m(x_1, x_3) \\ &+ m(x_2, x_3) + m(x_1, x_2, x_3) = 0.5, \end{aligned}$$

and

$$\begin{aligned} p_{\pi^5}(x_{\pi_2^5}) = p_{\pi^1}(x_1) &= \sum_{x_1 \in \mathcal{X}',\ x_3 \notin \mathcal{X}'} m(\mathcal{X}'), \\ &= m(x_1) + m(x_1, x_2) = 0.4, \end{aligned}$$

and $p_{\pi^5}(x_{\pi_3^5}) = p_{\pi^1}(x_2) = m(x_2) = 0.1$. It can be verified that the pair of permutations (π^1, π^3) and (π^5, π^6) give identical PMFs. Therefore, it can be concluded [11] that there are four PMFs associated to the belief function in Table I, that is $K(X) = Ch\{p_a, p_b, p_c, p_d\}$, where $p_a(x_1) = 0.4$, $p_a(x_2) = 0.1$, $p_a(x_3) = 0.5$, $p_b(x_1) = 0.6$, $p_b(x_2) = 0.35$, $p_b(x_3) = 0.05$ and $p_c(x_1) = 0.4$, $p_c(x_2) = 0.35$, $p_c(x_3) = 0.25$ and $p_d(x_1) = 0.6$, $p_d(x_2) = 0.1$, $p_d(x_3) = 0.3$ and where Ch denotes the convex hull. This means that $K(X)$ includes all PMFs such that:

$$p(x_i) = \alpha_1 p_a(x_i) + \alpha_2 p_b(x_i) + \alpha_3 p_c(x_i) + \alpha_4 p_d(x_i)$$

	Bel	Pl	m
$\{x_1\}$	0.4	0.6	0.4
$\{x_2\}$	0.1	0.45	0.1
$\{x_3\}$	0.05	0.3	0.05
$\{x_1, x_2\}$	0.5	0.5	0
$\{x_1, x_3\}$	0.65	0.65	0.2
$\{x_2, x_3\}$	0.4	0.4	0.25
$\{x_1, x_2, x_3\}$	1.0	1.0	0

TABLE I
EXAMPLE OF BELIEF FUNCTION

for $i = 1, 2, 3$ and for each $\alpha_j > 0$, $j = 1, 2, 3, 4$ such that $\sum_j \alpha_j = 1$. The fact that $K(X)$ is equivalent to the belief function in Table I can be verified by deriving the belief function from $K(X)$ using (5), i.e.:

$$Bel(\mathcal{X}') = \min_{p \in K(X)} \sum_{x_j} I_{\{\mathcal{X}'\}}(x_j) p(x_j),$$

for each $\mathcal{X}' \subseteq \mathcal{X}$. Notice that, in the minimization, it is sufficient to consider the extremes of $K(X)$, i.e., p_a, p_b, p_c, p_d, since the lower (upper) probability is obtained on the extremes. Thus $Bel(\cdot)$ gives the lower envelope of the closed-convex set of probability mass functions.

Given a belief function, the lower expected value of a function $h \in \mathcal{L}(\mathcal{X})$ can be calculated through (2), after having determined the set of PMFs associated to the belief function via (7). Consider, for instance, the belief function in Table I. Assuming that $h(x) = I_{\{x_1\}}(x) - I_{\{x_3\}}(x)$, i.e., $h(x_1) = 1$, $h(x_2) = 0$ and $h(x_3) = -1$, then it can be verified that $\underline{E}[h] = -0.1$ and $\overline{E}[h] = 0.55$. The lower expected value has been obtained w.r.t. the extreme PMF $p_a(\cdot)$, in fact:

$$\sum_{i=1}^{3} \left[I_{\{x_1\}}(x_i) - I_{\{x_3\}}(x_i) \right] p_a(x_i) = 0.4 - 0.5 = -0.1,$$

while the upper expectation w.r.t. $p_b(\cdot)$. For any other PMF in the convex hull of $\{p_a, p_b, p_c, p_d\}$ the expected value is included in $[-0.1, 0.55]$.

Belief functions are important special cases of CLPs. Beside this generality, the main difference between Walley's theory and theories based on belief functions (e.g., Dempster-Shafer and random set theory) is in the way evidence is combined and probabilities updated. Stemming from de Finetti's work on subjective probability, Walley gives a behavioural interpretation of upper and lower expectations in terms of buying and selling prices on gambles. According to this behavioural interpretation, the only consistent (coherent) way of defining a conditional model is through the Generalized Bayes Rule (4). He also shows that Dempster's rule of combination is incompatible with an approach based on coherence (it can incur in a sure loss and, thus, be unreasonable under a behavioural interpretation of probability [6, Ch. 5]).

B. Decision making

This section presents a brief discussion on decision making when uncertainty is represented by set of distributions. The

Bayesian methodology to decision making provides the action which maximises the expected value of some utility function. If $h_{a_j}(x)$ is the considered utility function, which depends on possible actions a_j and on the unknown value of the variable x, then a_i is preferred to a_j [12] if and only if

$$E_X(h_{a_j}) < E_X(h_{a_i}), \tag{8}$$

where $E_X(\cdot)$ denotes expectation w.r.t. x.

For example in classification, x is a discrete-valued class variable $x \in \{x_1, \ldots, x_n\}$ and action a_i represents the selection of class x_i. The utility function can be adopted to be the indicator function $h_{a_i}(x) = I_{\{x_i\}}(x)$, where $I_{\{x_i\}}(x) = 1$ if $x = x_i$ and zero otherwise. In this case $E_X(h_{a_i}) = \sum_j I_{\{x_i\}}(x_j)\, p(x_j) = p(x_i)$ and hence (8) becomes $p(x_j) < p(x_i)$. This means that one selects the class with the highest probability.

In imprecise probability one deals with interval-valued expectations, leading to the problem of decision making under imprecision [6]. A consequence of imprecision is that, when the lower and upper expectations are substantially different, then choosing a unique action may not be possible. Instead, a set of possible actions may be put forward to a decision maker. There are various ways for decision making with the convex set of probability functions, such as: interval dominance [6], maximality [6], minimax [12], Bayesian decision based on pignistic probabilities [2] etc.

C. Maximality

The maximality criterion was proposed by Walley [6]. Under maximality, a_i dominates (is preferred to) a_j if for all distributions p in the convex set $K(X)$, it holds that $E_X^p(h_{a_j}) < E_X^p(h_{a_i})$ or, equivalently, if

$$E_X^p(h_{a_i} - h_{a_j}) > 0 \;\; \forall p \in K(X), \tag{9}$$

where E_X^p denotes the expectation w.r.t. the density p. It can be proven that a necessary and sufficient condition for (9) to be satisfied is that

$$\underline{E}(h_{a_i} - h_{a_j}) > 0. \tag{10}$$

In the maximality criterion, actions are compared w.r.t. the same distribution, and thus a_i is said to dominate a_j if (9) is satisfied for each distribution in the convex set. This is a straightforward generalization of the Bayesian decision criterion (8) to set of distributions. For this reason maximality is the decision criterion adopted in this paper.

Consider again the example in Table I and assume that the elements x_i in $\mathcal{X} = \{x_1, x_2, x_3\}$ denote possible classes in a a classification problem. Hence, given the bba m, we aim to decide which is the most probable class. In this respect, we can compute the pignistic transformation [2]:

$$BetP(x) = \sum_{x \in \mathcal{X}' \subseteq \mathcal{X}} \frac{m(\mathcal{X}')}{|\mathcal{X}'|}. \tag{11}$$

and, then, use $BetP$ in (8) for decision making. The pignistic probability for the example in Table I is:

$$BetP(x_1) = 0.5,\;\; BetP(x_2) = 0.225,\;\; BetP(x_3) = 0.275$$

and, thus, the most probable class is x_1 (obtained from (8) in the case $h_{a_j} = I_{\{x_j\}}$). However, we could also use the maximality criterion for decision making. In the case $h_{a_i}(x) - h_{a_j}(x) = I_{\{x_i\}}(x) - I_{\{x_j\}}(x)$, it follows that

$$\underline{E}[I_{\{x_1\}} - I_{\{x_2\}}] = 0.05,\;\; \underline{E}[I_{\{x_1\}} - I_{\{x_3\}}] = -0.1,$$
$$\underline{E}[I_{\{x_2\}} - I_{\{x_3\}}] = -0.4$$

where the lower expectations have been obtained considering the extreme distributions $\{p_a, p_b, p_c, p_d\}$ computed in Section II-A. Since only $\underline{E}[I_{\{x_1\}} - I_{\{x_2\}}] > 0$ this implies that the undominated classes are $\{x_1, x_3\}$. Notice in fact that dominance is an anti-commutative and transitive operation. Thus, under maximality, we can just state that the most probable class belongs to the set $\{x_1, x_3\}$. This happens because the set $K(X)$ includes these two extreme PMFs: $p_a(x_1) = 0.6$, $p_a(x_2) = 0.35$, $p_a(x_3) = 0.05$ and $p_b(x_1) = 0.4$, $p_b(x_2) = 0.1$, $p_b(x_3) = 0.5$. We can see that according to p_a, x_1 is more probable than x_3, but according to p_b, x_3 is more probable than x_1. Thus, we cannot decide which is the most probable between x_1 and x_3. Maximality provides a more cautious and, thus, more robust decision criterion than that based on the pignistic transformation.

III. Problem description

Consider two domains $\mathcal{X} \subseteq \mathbb{N}$ is a discrete set of (target) classes, and $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$ is the target feature domain ($n_z \geq 1$, that is $z \in \mathcal{Z}$ can be a multidimensional feature vector). The inference is carried out on $\mathcal{X}$, which is hidden, being not directly observable. The features are the measurements available for making inference; they are related to the hidden target classes via the likelihood function.

In some cases the likelihood function is precisely known. For example, let $x \in \mathcal{X}$ be a class and $z = H(x) + v$ the measured feature, where $H(\cdot)$ is a known possibly nonlinear function and v is the measurement noise, distributed according to the probability density function $p_v(\cdot)$. The likelihood function in this case is precisely known, i.e. $p(z|x) = p_v(z - H(x))$.

In the following, however, we are interested in a more realistic case of imprecise and vague likelihood functions. The features in this case need to be modelled by random sets, rather than random variables. The particular examples of random sets are intervals, fuzzy membership functions, Dempster-Shafer (DS) basic belief assignments (bba's) and fuzzy DS bba's. In order to fix the ideas, let us consider an example of imprecise likelihood functions, taken from [1].

Suppose there are only three classes of aircraft, i.e. $\mathcal{X} = \{1, 2, 3\}$, where class 1 are commercial planes, class 2 are large military aircraft and class 3 are light and agile military aircraft. The available target feature for classification is the maximum observed acceleration (during a certain interval of time). This feature is related to classes as follows: for class 1, the acceleration is rarely higher than $1g$ ($g = 9.81$ m/s^2 is the acceleration due to gravity); targets of class 2 sometimes perform mild evasive manoeuvres, but their maximum acceleration (due to their size) cannot be higher than $4g$; targets of class 3 are light and agile, with highly trained pilots, so the maximum acceleration can go up to $7g$. The steady-state of acceleration, however, for all three classes of targets can be considered to be zero (because of minimal fuel consumption, minimal stress for pilots, etc.). Finally, the knowledge of prior class probabilities is unavailable.

IV. Modelling

Based on the above description of the problem, consider two cases of the feature domain:

- the discrete (finite) domain of accelerations, $\mathcal{Z} = \{z_1, z_2, z_3\}$, where z_1 corresponds to small acceleration, i.e. $z_1 = \{z : |z| \leq 1g\}$, $z_2 = \{z : 1g < |z| \leq 4g\}$ represents moderate acceleration and $z_3 = \{z : 4g < |z| \leq 7g\}$ high acceleration;
- the continuous domain of acceleration, $\mathcal{Z} = (-\infty, \infty)$.

The likelihood of measurement $z \in \mathcal{Z}$ is referred to as imprecise because there is no unique mapping from $\mathcal{X}$ to $\mathcal{Z}$. Take for example the discrete domain $\mathcal{Z} = \{z_1, z_2, z_3\}$: class 2 maps into a *set* $\{z_1, z_2\}$, while class 3 maps into a set $\{z_1, z_2, z_3\}$. We can model this uncertainty by using set of likelihoods. For instance, the set of likelihoods $K(Z|x_3)$ that maps class 3 into the set $\{z_1, z_2, z_3\}$ may be defined as:

$$K(Z|x_3) = \{p(\cdot|x_3) : \; p(z_1|x_3) + p(z_2|x_3) + p(z_3|x_3) = 1\},$$

i.e., $K(Z|x_3)$ includes all the possible conditional PMFs. This is the least uninformative CLP that models the probabilistic relationship between acceleration and class 3. The correspondence between $\mathcal{X}$ and $\mathcal{Z}$ may also be specified by a more informative set of PMFs. Let Σ_x denote a random set representation of the feature, specified by a bba. Then Σ_x can be expressed as:

$$\Sigma_x = \{(m, A) : A \subseteq \mathcal{Z}; m > 0; \sum_{m:(m,A)\in\Sigma_x} m(A) = 1\} \tag{12}$$

where m represents the bba associated to $Bel(\cdot|x)$. A possible specification of Σ_x for each class in the example above is:

$$\Sigma_{x_1} = \Big\{(1, \{z_1\})\Big\} \tag{13}$$

$$\Sigma_{x_2} = \Big\{(0.6 - \frac{\epsilon}{2}, \{z_1\}), (\epsilon, \{z_2\}), (0.4 - \frac{\epsilon}{2}, \{z_1, z_2\})\Big\} \tag{14}$$

$$\Sigma_{x_3} = \Big\{(0.5 - \frac{\epsilon}{3}, \{z_1\}), (\epsilon, \{z_3\}), (0.3 - \frac{\epsilon}{3}, \{z_1, z_2\}), (0.2 - \frac{\epsilon}{3}, \{z_1, z_2, z_3\})\Big\}. \tag{15}$$

where $\epsilon > 0$ is an arbitrarily small number which has been introduced to allow the applicability of GBR (4), i.e., $p(z) > 0$ for all distributions in $K(Z|X)$.

The first bba in (13) is categorical, with a singleton focal element; it describes a precise likelihood, i.e., $p(z_1|x_1) = 1$. The other two bba's, however, are multi-focal and can be represented by:

$$K(Z|x_2) = Ch\{p_1, p_2\}, \tag{16}$$

with $p_1(z_1|x_2) = 1 - \epsilon$, $p_1(z_2|x_2) = \epsilon$ and $p_2(z_1|x_2) = 0.6 - \frac{\epsilon}{2}$, $p_2(z_2|x_2) = 0.4 + \frac{\epsilon}{2}$, while

$$K(Z|x_3) = Ch\{p_1, p_2, p_3\}, \tag{17}$$

with $p_1(z_1|x_3) = 1 - \epsilon$, $p_1(z_3|x_3) = \epsilon$, $p_2(z_1|x_3) = 0.5 - \frac{\epsilon}{3}$, $p_2(z_2|x_3) = 0.5 - \frac{2\epsilon}{3}$, $p_2(z_3|x_3) = \epsilon$, $p_3(z_1|x_3) = 0.5 - \frac{\epsilon}{3}$, $p_3(z_2|x_3) = 0.3 - \frac{\epsilon}{3}$, $p_3(z_3|x_3) = 0.2 + \frac{2\epsilon}{3}$, and $p_4(z_1|x_3) = 0.8 - \frac{2\epsilon}{3}$, $p_4(z_2|x_3) = 0$, $p_4(z_3|x_3) = 0.2 + \frac{2\epsilon}{3}$. For the continuous measurement domain, a random set Σ_x is specified by a fuzzy DS bba as follows:

$$\Sigma_x = \{(w, \mu) : \mu \in \mathfrak{U}(\mathcal{Z}); \; w > 0; \sum_{w:(w,\mu)\in\Sigma_x} w = 1\}, \tag{18}$$

where $\mathfrak{U}(\mathcal{Z})$ is the set of fuzzy membership functions defined on $\mathcal{Z}$. One possibility for the specification of Σ_x in accordance with the problem description is:

$$\Sigma_{x_i} = \{(w_i, \mu_i)\} \tag{19}$$

for $i = 1, 2, 3$, where

$$w_i = 1, \quad \mu_i(z) = \exp\Big\{-\frac{z^2}{2\sigma_i^2}\Big\}, \tag{20}$$

and $\sigma_1 = 0.4g$, $\sigma_2 = 1.6g$ and $\sigma_3 = 2.8g$. These values of standard deviations were adopted to ensure that $Pr\{|z| < \gamma\} = 0.99876$, for $\gamma = 1g$, $4g$, and $7g$, representing the limits for class 1, 2, and 3, respectively. All three fuzzy DS bba's of (19) have a single focal element. Other choices are possible too. For example, a fuzzy version of the bba in (13)–(15) can be specified as:

$$\Sigma_{x_1} = \{(1, \mu_1)\}, \tag{21}$$

$$\Sigma_{x_2} = \{(0.6, \mu_1), (0.4, \mu_2)\} \tag{22}$$

$$\Sigma_{x_3} = \{(0.5, \mu_1), (0.3, \mu_2), (0.2, \mu_3)\} \tag{23}$$

where μ_i were defined in (20).

Within the CLP framework, the continuous case is taken into account by considering set of densities instead of set of probability mass functions. There are infinite ways to define a set of densities, hereafter we just consider the so-called lower and upper bounded density model. This model is defined as follows:

$$K(Z|x_i) = \{p(z|x_i) : \; l(z|x_i) \leq p(z|x_i) \leq u(z|x_i)\} \tag{24}$$

where the lower l and upper u densities are unnormalized (and possibly improper) densities. Thus, $K(Z|x_i)$ includes all the densities lower and upper bounded by l and, respectively, u. For the classification problem, we will consider $u(z|x_i) = \sum_i w_i \cdot \mu_i(z)$ and $l(z|x_i) = \epsilon_i u(z|x_i)$, with $\epsilon_i \in (0, 1)$ and w_i, μ_i defined in (19)–(23).

V. Classifier based on TBM

As before, $x \in \mathcal{X}$ is a class and $z \in \mathcal{Z}$ is a feature measurement. Based on TBM framework [1], the classifier for imprecise likelihoods is obtained as follows. Starting from precise likelihoods $p(z|x_i)$ (same that a Bayesian classifier would use), it treats them as being pignistic function. The first step is to build the least committed plausibility function (defined in (27)) over the observation domain which corresponds to the pignistic function, followed by the application of the Generalised Bayesian Theorem [1]:

$$m(\mathcal{X}'|z) \quad \propto \quad \prod_{x_i \in \mathcal{X}'} pl(z|x_i) \prod_{x_i \in \mathcal{X} \setminus \mathcal{X}'} (1 - pl(z|x_i)) \quad (25)$$

for each $\emptyset \neq \mathcal{X}' \subseteq \mathcal{X}$.[5] Finally the last step is to normalize $m(\mathcal{X}'|z)$ and then to apply the pignistic transformation (11) to compute the pignistic probabilities $BetP(x_i|z)$ for each x_i.

Concerning the classification problem of Section IV, the least-committed plausibility function relative to the Gaussian density $p(z|x_i) = \mathcal{N}(z; 0, \sigma_i^2)$ can be obtained as follows [1]:

$$\mu_i^{LC}(z) \quad = \quad \frac{2y}{\sqrt{2\pi}} e^{-y^2/2} + \text{erfc}(y/\sqrt{2}) \qquad (26)$$

where $\text{erfc}(s) = \frac{2}{\sqrt{\pi}} \int_s^\infty e^{-t^2} dt$ and $y = z/\sigma_i$. Taking account the fuzzy DS bba defined by (18), the plausibility function to be used in (25) for the classification problem is:

$$pl(z|x_i) = \sum_{(w,\mu) \in \Sigma_x} w \cdot \mu^{LC}(z). \qquad (27)$$

VI. Classifier based on RS approach

The most used approach for classification in the realm of RS theory is as follows [4, Ch.6]. The Bayes classifier for imprecise likelihoods[6] is given by:

$$P(x|z) = \frac{g(z|x) \cdot P(x)}{\sum_{x \in \mathcal{X}} g(z|x) \cdot P(x)} \qquad (28)$$

where $P(x)$ is the prior class probability and $g(z|x)$ is the *generalized likelihood function*, defined as:

$$g(z|x) = Pr\{z \in \Sigma_x\}. \qquad (29)$$

In the DS theory, the probability on the r.h.s. of (29) is referred to as the plausibility function on singletons, i.e. $g(z|x) \equiv pl_{\Sigma_x}^{\mathcal{Z}}(z)$. Assuming first that the domain $\mathcal{Z}$ is discrete, the plausibility function on singletons corresponding to the random set Σ_x specified in (12), is given by:

$$pl_{\Sigma_x}^{\mathcal{Z}}(z) = \sum_{m:(m,B) \in \Sigma_x,\ \{z\} \cap B \neq \emptyset} m(B) \qquad (30)$$

For the continuous domain $\mathcal{Z}$ and a fuzzy DS bba defined by (18), the plausibility function on singletons is defined as:

$$pl_{\Sigma_x}^{\mathcal{Z}}(z) = \sum_{(w,\mu) \in \Sigma_x} w \cdot \mu(z). \qquad (31)$$

In summary, based on RS approach, the Bayes classifier for imprecise likelihoods, where the correspondence between $\mathcal{X}$ and $\mathcal{Z}$ is specified by a random closed set Σ_x, is carried out in two steps. The first step is to compute the plausibility function (on singletons) of Σ_x. The second step is to apply the Bayes rule (28).

[5]This has been obtained assuming a vacuous prior on $\mathcal{X}$.

[6]The same form is applicable to any non-traditional data represented by random sets.

The fact that (28) depends only on the values of the plausibility function on singletons is an advantage from a computational point of view but, on the other hand, implies a large loss of information for RS approach.[7] In fact, there can be RSs with a very different information content but with same plausibility on (some) singletons. For instance, $m_1(x_1) = 0.5$, $m_1(x_2) = 0.4$, $m_1(x_3) = 0.1$ and $m_2(x_1) = 0.3$, $m_2(x_2) = 0.4$, $m_2(x_3) = 0.1$, $m_2(x_1, x_3) = 0.2$ have same plausibility on the singletons x_1 and x_2, but different information content (i.e., m_1 is a standard PMF). A similar issue arises in TBM approach, since also the pignistc transformation implies a loss of information w.r.t. the original belief function.

VII. Classifier based on imprecise probability

In the CLP framework, updating is performed through GBR. Thus, given the conditional CLP $\underline{E}_Z(\cdot|X)$ and prior CLP $\underline{E}_X$, first we compute the joint CLP $\underline{E}_{X,Z}[h] = \underline{E}_X[\underline{E}_Z[h|X]]$ via marginal extension (3) and then we use GBR to compute the posterior CLP $\underline{E}_X(\cdot|Z)$. In particular, for the classification likelihood in Section IV, if $p(X)$ is assumed to be a precise probability, the posterior CLP can be computed as follows:

$$\underline{E}_X[f|z] = \min_{p(\cdot|X) \in K(Z|X)} \frac{\sum_{x \in \mathcal{X}} f(x) p(z|x) p(x)}{\sum_{x \in \mathcal{X}} p(z|x) p(x)} \qquad (32)$$

In practice, in (32) is sufficient to consider the extremes of the various sets $K(Z|x_i)$. The lower expectation $\underline{E}_X(\cdot|z)$ is then employed for decision making based on the maximality criterion defined in Section II-C. One of the main differences w.r.t. TBM and RS approach is that the result of (32) is not a standard expectation w.r.t. a single PMF, but a CLP associated to a set of PMFs.[8].

VIII. Numerical results

In this section, we compare TBM, RS and CLP based approach in the classification problem described in Section IV.

a) Discrete feature domain: The set of classes is $\mathcal{X} = \{x_1, x_2, x_3\}$, the feature domain is $\mathcal{Z} = \{z_1, z_2, z_3\}$. Considering the RS approach, the generalised likelihood function $g(z_i|x_j)$, for $i, j = 1, 2, 3$, is the plausibility function on singletons of random set Σ_j specified in (13)–(15). Using (30) it follows that:

$$g(z_1|x_1) = g(z_1|x_2) = g(z_1|x_3) = 1 - \epsilon,$$
$$g(z_2|x_1) = 0,\ g(z_2|x_2) = 0.4 + \frac{\epsilon}{2},\ g(z_2|x_3) = 0.5 + \frac{2\epsilon}{3},$$
$$g(z_3|x_1) = g(z_3|x_2) = 0,\ g(z_3|x_3) = 0.2 + \frac{2\epsilon}{3}.$$

[7]A solution to solve this issue is proposed in [4, Ch. 7] employing RS over set of likelihoods.

[8]There is no loss of information, since the functional $\underline{E}_X(\cdot|z)$ has the same information content of $K(X|Z)$. That is, at the changing of f in (32) the PMF which obtains the minimum can change, while in the TBM and RS approach the PMF is always the same, i.e., the pignistic probability or the probability in (28).

Assuming uniform priors on classes, i.e. $p(x_1) = p(x_2) = p(x_3) = 1/3$, it follows from (28) that:

$$p(x_1|z_1) = p(x_2|z_1) = p(x_3|z_1) = 1/3. \tag{33}$$

The same result is obtained via TBM approach. It appears that the feature measurement z_1 (small acceleration) does not alter the priors (it is uninformative). This is perfectly in accordance with our intuition because all three classes of aircraft fly (most of the time) with small acceleration.

Considering a CLP based approach and $f = I_{\{x_i\}} - I_{\{x_j\}}$ in (32), we can compute the dominance condition (10), i.e., $\underline{E}_X(I_{\{x_i\}} - I_{\{x_j\}}|z_1) > 0$, which reduces to:

$$\min_{p(\cdot|X)\in K(z_1|X)} \frac{p(z_1|x_i) - p(z_1|x_j)}{p(z_1|x_1) + p(z_1|x_2) + p(z_1|x_3)} \tag{34}$$

where the prior $p(x_i) = 1/3$ has been simplified from numerator and denominator. Since for maximality we are just interesting on the sign of $\underline{E}_X(I_{\{x_1\}} - I_{\{x_2\}}|z_1)$, we can neglect the contribution of the denominator. Thus, x_i dominates x_j if

$$\min_{p(\cdot|X)\in K(z_1|X)} p(z_1|x_i) - p(z_1|x_j) > 0. \tag{35}$$

From (16)–(17), it follows that $p(z_1|x_1) = 1 - \epsilon$, $p(z_1|x_2) \in [0.6 - \epsilon/2, 1 - \epsilon]$, $p(z_1|x_3) \in [0.5 - \epsilon/3, 1 - \epsilon]$. Hence, we conclude that (35) is non positive for all comparisons i, j and, thus, that all the classes $\{x_1, x_2, x_3\}$ are undominated. This is in perfect agreement with (33), we cannot take a decision.

If the feature measurement is z_2 (medium acceleration), based on the RS approach one gets (neglecting the contribution of ϵ, which has been assumed to be small):

$$p(x_1|z_2) = 0,\ p(x_2|z_2) = 4/9,\ p(x_3|z_2) = 5/9. \tag{36}$$

The pignistic probabilty obtained through TBM is $BetP(x_1|z_2) = 0$, $BetP(x_2|z_2) \approx 0.43$, $BetP(x_3|z_2) \approx 0.57$ which is numerically close to the RS approach. Using the Voorbraak approximation [13], instead of the pignistic transform, leads to identical results to those using the RS approach (as claimed in [4]).

Under CLP and maximality, since $p(z_2|x_1) = 0$, $p(z_2|x_2) \in [\epsilon, 0.4 + \epsilon/2]$, $p(z_2|x_3) \in [\epsilon, 0.5 + 2\epsilon/3]$, we can conclude that $\{x_2, x_3\}$ are undominated. This result seems to be more in agreement with our intuition than that derived via TBM and RS approach (because TBM and RS approaches slightly favour class 3 over class 2).

Finally, based on the RS and TBM approach, measurement z_3 results in classification probabilities:

$$p(x_1|z_3) = p(x_2|z_3) = 0, \quad p(x_3|z_3) = 1. \tag{37}$$

Under CLP and maximality, since $p(z_3|x_1) = 0$, $p(z_3|x_2) = 0$, $p(z_3|x_3) \in [\epsilon, 0.2 + 2\epsilon/3]$, we conclude that 3 is the only undominated class. This is in accordance with TBM and RS approaches and our intuition.

b) Continuous feature domain.: Again $\mathcal{X} = \{1, 2, 3\}$ but this time $\mathcal{Z} = (-\infty, \infty)$. Let us first work out the case where Σ_j, $j = 1, 2, 3$ is specified by (19)–(20). According to (31), in this case the generalised likelihood function is simply:

$$g(z|x_j) = \mu_j(z) \tag{38}$$

for $z \in \mathcal{Z}$ and $j = 1, 2, 3$. Application of (28) results in classification probabilities depicted in Figure 1.(a). This figure is practically the same of Figure 6 in [1], which was obtained using the TBM machinery on the continuous feature domain with the plausibility (26). For very small values of acceleration z, all three classes are equally probable. It was argued in [1] that this is again perfectly in accordance with our intuition (as opposed to the classification results obtained by the standard Bayesian classifier, which uses $\mathcal{N}(z; 0, \sigma_i^2)$ as likelihood, reported in Figure 2). Finally, for the case of

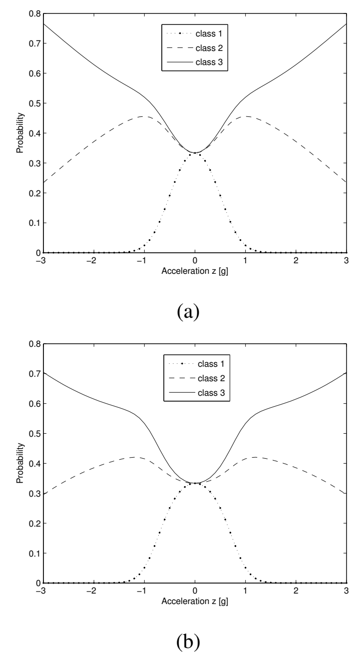


Fig. 1. Posterior class probabilities $Pr\{x_j|z\}$, $j = 1, 2, 3$ for Σ_j defined by (a) equations (19)–(20); (b) equations (21)–(23).

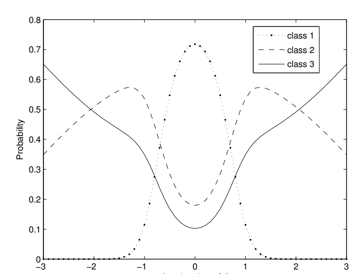


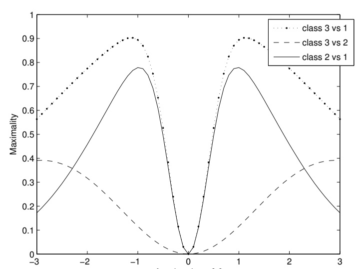
Fig. 2. Posterior class probabilities for a Bayesian classifier based on the likelihoods $\mathcal{N}(z; 0, \sigma_i^2)$ and uniform prior.

imprecise likelihoods where the feature is specified by random set Σ_j, $j = 1, 2, 3$ in equations (21)–(23), Figure 1.(b) depicts the classification probabilities. The result is very similar to that of Figure 1.(a) and the same arguments apply.

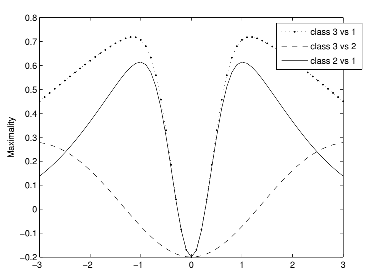
Consider now the CLPs framework and the set of densities in (24), we can compute the dominance condition (10) for $f = I_{\{x_i\}} - I_{\{x_j\}}$ which, as in the discrete case, reduces to:

$$\min_{p(\cdot|X) \in K(Z|X)} p(z|x_i) - p(z|x_j) > 0,$$

or, equivalently, to $\epsilon_i u(z|x_i) - u(z|x_j) > 0$. In other words, x_i dominates x_j for the values of z such that $\epsilon_i u(z|x_i) - u(z|x_j) > 0$. In the case $\epsilon_i = \epsilon \approx 1$ for all $i = 1, 2, 3$ and $w_i, \mu_i(z)$ defined as in (19)–(20), the plots of $u(z|x_i) - u(z|x_j)$, labelled as "class i vs. j"[9], are shown in Figure 3.(a). It can be noticed that all the plots are always positive. This means that class i dominates class j for all values of $z \in \mathcal{Z} \backslash \{0\}$. In this case, the classification results coincide with those depicted in Figures 1.(a), i.e., class 3 is more probable than class 2 which, in turn, is more probable than class 1 for each $z \neq 0$. Same happens in the case $w_i, \mu_i(z)$ are defined as in (21)–(23). At the decreasing of ϵ, TBM and RS approaches still provide the same results, since the plausibility (upper density) does not change. Conversely, according to the CLP framework, the classification results become more uncertain. Figure 3.(b) shows the case $\epsilon = 0.8$. Since $\epsilon u(z|x_i) - u(z|x_j)$



(a)



(b)

Fig. 3. Results of classification with CLP for $w_i, \mu_i(z)$ defined as in (19)–(20) (a) case $\epsilon = 0$; (b) $\epsilon = 0.8$.

is negative in $z \in [-0.28, 0.28]$ for each i, j[10], it follows that in $z \in [-0.28, 0.28]$ all classes $\{1, 2, 3\}$ are undominated. For $z \in [-1.3, -0.28] \cup [0.28, 1.3]$, classes $\{2, 3\}$ are undominated. This means that if the acceleration is between $[-0.28, 0.28]$ we cannot decide which is the class of the aircraft; if the acceleration is in $[-1.3, -0.28] \cup [0.28, 1.3]$ we can say that the aircraft is not of class 1; if it is greater than $|1.3|$ we can say that the aircraft is of class 3.

[9] In this Figure, we have compared only 3 vs. 1, 3 vs 2 and 2 vs. 1, since for the opposite comparisons $u(z|x_i) - u(z|x_j)$ is clearly always negative.

[10] As before, we have compared only 3 vs. 1, 3 vs 2 and 2 vs. 1, since for the opposite comparisons i cannot dominate j since $u(z|x_i) - u(z|x_j) < 0$.

IX. CONCLUSIONS

In many classification problems it is impossible to specify the precise model of the feature measurement likelihood function. It is known that the standard Bayes classifier in those circumstances is inappropriate to use. The paper highlights the differences between three suitable solutions, based on (1) the Transferrable Belief model (TBM), (2) the random set framework for Bayesian classification and (3) the imprecise probability approach. A simple target classification based on acceleration feature is used for comparison. The results illustrate that the RS approach produces identical results to those obtained using the TBM classifier, if equivalent measurement models are employed. The imprecise probability theory provides a more general framework for reasoning under uncertainty. Using the maximality criterion for decision making the paper showed how to obtain similar results to those produced by the TBM or RS classifier.

ACKNOWLEDGEMENTS

This work has been partially supported by the Swiss NSF grants n. 200020-121785/1.

REFERENCES

[1] B. Ristic and P. Smets, "Target classification approach based on the belief function theory," *IEEE Trans. Aerospace and Electronic Systems*, vol. 41, pp. 574–583, April 2005.

[2] P. Smets, "Data fusion in the transferable belief model," in *Proc. 3rd Intern. Conf. Information Fusion (FUSION 2000)*, (Paris, France), pp. PS21–33, 2000.

[3] P. Smets, "Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem," *International Journal of Approximate Reasoning*, vol. 9, pp. 1–35, 1993.

[4] R. Mahler, *Statistical Multisource Multitarget Information Fusion*. Artech House, 2007.

[5] R. Mahler, "Bayesian fusion and filtering of nontraditional data." Tutorial notes, 13th Int. Conf. Information Fusion, Edinburgh, UK, July 2010.

[6] P. Walley, *Statistical Reasoning with Imprecise Probabilities*. New York: Chapman and Hall, 1991.

[7] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.

[8] D. Dubois and H. Prade, *Possibility Theory*. New York: Plenum Press, 1988.

[9] A. Benavoli, M. Zaffalon, and E. Miranda, "Robust filtering through coherent lower previsions," *accepted on IEEE Trans. on Automatic Control, doi:10.1109/TAC.2010.2090707*, August, 2011.

[10] B. Ristic, "Target classification with imprecise likelihoods: Mahler's approach," *IEEE Trans. Aerospace and Electronic Systems*, vol. 47, April 2011.

[11] F. Cuzzolin, "Credal semantics of Bayesian transformations," in *Proc. ISIPTA'09*, (Durham, UK), pp. 129–138, July 2009.

[12] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York: Springer Series in Statistics, 1985.

[13] F. Voorbraak, "A computationally efficient approximation of Dempster-Shafer theory," *International Journal of Man-Machine Studies*, vol. 30, no. 5, pp. 525–536, 1989.