A Hybrid Belief Rule-Based Classification System Based on Uncertain Training Data and Expert Knowledge

Lianmeng Jiao, Thierry Denœux, *Member, IEEE* and Quan Pan, *Member, IEEE*

Abstract

In some real-world classification applications, such as target recognition, both training data collected by sensors and expert knowledge may be available. These two types of information are usually independent and complementary, and both are useful for classification. In this paper, a hybrid belief rule-based classification system (HBRBCS) is developed to make joint use of these two types of information. The belief rule structure, which is capable of capturing fuzzy, imprecise, and incomplete causal relationships, is used as the common representation model. With the belief rule structure, a data-driven belief rule base (DBRB) and a knowledge-driven belief rule base (KBRB) are learnt from uncertain training data and expert knowledge, respectively. A fusion algorithm is proposed to combine the DBRB and KBRB to obtain an optimal hybrid belief rule base (HBRB). A belief reasoning & decision making module is then developed to classify a query pattern based on the generated HBRB. An airborne target classification problem in the air surveillance system is studied to demonstrate the performance of the proposed HBRBCS for combining both uncertain sensor measurements and expert knowledge to make classification.

Index Terms

Hybrid classification system, belief rule, uncertain data, expert knowledge.

I. INTRODUCTION

Pattern classification is the study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decisions about the categories of the patterns [1]. According to the type of information used in modeling, pattern classification methods can be categorized into data-driven and knowledge-driven [2]. Data-driven models are based on the learning of training data characterizing the

problem under modeling. The data is usually collected by the sensors observing the environment [3]. Popular data-driven models include k-nearest neighbor rule, decision trees, support vector machines, artificial neural networks, and fuzzy rule-based systems, etc. [4]. In contrast, knowledge-driven models are based on expert knowledge understanding a particular domain or problem. The expert knowledge is often coded into rules or relations, based on which inference is made [5]. Examples of the most popular systems using knowledge-driven models are expert systems [6] and decision support systems [7].

In some real-world classification applications, both training data and expert knowledge may be available. For example, for target recognition problem [8], [9], the historical measurements from a long period of collection by sensors can be used to train the target recognition system. In addition, expert knowledge about target characteristics obtained from the manufacturers or intelligence also provides important information for target recognition. In the classification process, training data and expert knowledge are complementary [10]. The training data tends to provide a relatively fine estimate for the real class-conditioned distribution but may be unreliable in some specific regions of feature space due to limited training patterns and the potential measurement noise, whereas the expert knowledge usually provides a relatively rough but overall reliable estimate for the real class-conditioned distribution. In the past, several methods have been proposed to address the classification problem based on both training data and expert knowledge. For instance, Zhou et al. [11] proposed to build an initial rule base from the available expert knowledge, and then to optimize the rule base by adding or pruning rules based on training data. In [12], Tang et al. developed a knowledge-based naive Bayes classifier, which uses training data to estimate the involved conditional probabilities. Later, in [13], they proposed another similar method, which builds a fuzzy rule-based system based on expert knowledge and then uses training data to optimize the involved fuzzy membership functions. In essence, these methods follow the same idea that first building a base model from expert knowledge and then optimizing this model based on training data. However, one major disadvantage of this idea is that the weights of training data and expert knowledge cannot be adjusted according to the qualities of these two types of information. In this paper, we intend to address the hybrid classification problem with a different idea. That is, to build a data-driven model and a knowledge-driven model independently, and then to combine them into an adaptive hybrid classification model by taking into account their weights.

In order to combine training data and expert knowledge for classification, a common representation model that can make use of both types of information is needed. The IF-THEN rule is a good representation model, because, on the one hand, the IF-THEN rule can be learnt from training data and, on the other hand, expert knowledge is also easily coded into IF-THEN rule. Many fuzzy rule-based systems have been proposed to deal with the classification problems using either training data or expert knowledge [14], [15], [16], [17]. However, different types of uncertainty may coexist in real-world applications, e.g., both training data and expert knowledge may be imprecise or incomplete. The fuzzy rule-based systems, which are based on fuzzy set theory [18], cannot address the imprecise or incomplete information effectively in the modeling and reasoning processes. The belief function theory, also called Dempster-Shafer theory, proposed and developed by Dempster [19], Shafer [20] et al., has become one of the most powerful frameworks for uncertainty modeling and reasoning. Many researchers have investigated the

relationship between fuzzy set theory and belief function theory and suggested different ways of integrating them. Among them, Yang et al. [21] extended the fuzzy rule with belief function theory and proposed a new knowledge representation scheme in a belief rule structure, which is capable of capturing fuzzy, imprecise, and incomplete causal relationships. The belief rule structure has been successfully applied in a number of applications, such as clinical risk assessment [22], inventory control [23], and fault diagnosis [24].

In our previous work [25], a belief rule-based classification system was developed based on uncertain training data, which shows great advantages of the belief rule for modeling uncertain information in complex classification problems. In this paper, we focus on more general cases, where both training data and expert knowledge are available. Based on the belief rule structure, a hybrid belief rule-based classification system (HBRBCS) is developed to make good use of these two types of information. The proposed HBRBCS is composed of two main components: a hybrid belief rule base (HBRB) that establishes the association between the feature space and the class space, and a belief reasoning & decision making module that provides a mechanism to classify a query pattern based on the HBRB. With the belief rule structure, a data-driven belief rule base (DBRB) and a knowledge-driven belief rule base (KBRB) are learnt from uncertain training data and expert knowledge, respectively. A fusion algorithm is proposed to combine the DBRB and KBRB to obtain an optimal HBRB. Based on the generated HBRB, a belief reasoning algorithm is applied to classify a query pattern in a robust way and both a hard and soft decision strategies are provided for user to select depending on different needs. At last, the proposed HBRBCS is applied to solve an airborne target classification problem based on uncertain sensor measurements and expert knowledge.

The rest of this paper is organized as follows. In Section II, the basics of belief function theory and the belief rule structure are introduced. The hybrid belief rule-based classification system is developed in Section III, and then an airborne target classification problem in the air surveillance system is studied to demonstrate the performance of the proposed HBRBCS for combining both uncertain training data and expert knowledge to make classification in Section IV. Finally, Section V concludes the paper.

II. THEORETICAL BACKGROUND

A. Basics of Belief Function Theory

Let $\Theta = \{\theta_1, \theta_2, \cdots, \theta_n\}$ be a finite set of mutually exclusive and exhaustive hypotheses about the problem domain called the *frame of discernment*. A *basic belief assignment* (BBA) is a function m: $2^\Theta \to [0, 1]$, verifying

$$\mathrm{m}(\emptyset) = 0 \quad \text{and} \quad \sum_{A \in 2^\Theta} \mathrm{m}(A) = 1. \tag{1}$$

The quantity m(A) measures the degree of belief exactly assigned to proposition A, given the available evidence. A subset $A \in 2^\Theta$ with m$(A) > 0$ is called a *focal element*. The belief assigned to Θ represents the degree of *global ignorance*. A BBA m is said to be *vacuous* if there is full global ignorance.

Two evidential functions derived from the BBA are the *belief function* and *plausibility function* defined as

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} \mathrm{m}(B) \quad \text{and} \quad \mathrm{Pl}(A) = \sum_{B \cap A \neq \emptyset} \mathrm{m}(B). \tag{2}$$

The belief $\text{Bel}(A)$ and the plausibility $\text{Pl}(A)$ are usually interpreted as lower and upper probabilities of proposition A. For decision making, the *pignistic probability* $\text{BetP}(A)$ [26] is usually used to approximate the unknown probability in $[\text{Bel}(A), \text{Pl}(A)]$, defined as

$$\text{BetP}(A) = \sum_{B \subseteq \Theta;\ A \cap B \neq \emptyset} \frac{|A \cap B|}{|B|} \text{m}(B), \tag{3}$$

where $|X|$ is the cardinality of set X.

For the case of partially reliable source of evidence characterized by BBA m, Shafer's discounting operation with reliability factor $\alpha \in [0, 1]$ is defined as

$$^{\alpha}\text{m}(A) = \begin{cases} \alpha \text{m}(A), & \text{for } A \neq \Theta \\ \alpha \text{m}(\Theta) + (1 - \alpha), & \text{for } A = \Theta. \end{cases} \tag{4}$$

Consider now two distinct pieces of evidence on the same frame of discernment Θ represented by two BBAs m_1 and m_2. They can be combined using Dempster's rule to generate a new BBA $\text{m}_1 \oplus \text{m}_2$ as follows

$$\text{m}_1 \oplus \text{m}_2(A) = \begin{cases} 0, & \text{for } A = \emptyset \\ \dfrac{\sum\limits_{B \cap C = A} \text{m}_1(B)\text{m}_2(C)}{1 - \sum\limits_{B \cap C = \emptyset} \text{m}_1(B)\text{m}_2(C)}, & \text{for } A \in 2^{\Theta} \text{ and } A \neq \emptyset. \end{cases} \tag{5}$$

This rule is commutative, associative, and it has the vacuous mass function $\text{m}(\Omega) = 1$ as the unique neutral element.

B. Belief Rule Structure

In order to capture the uncertain association between the feature space and the consequence space, Yang et al. [21] extended the fuzzy rule with belief function theory and proposed a new knowledge representation scheme in a belief rule structure as follows.

> Belief Rule R^q : If x_1 is A_1^q and x_2 is A_2^q and $\cdots$ and x_P is A_P^q, then the consequence is
> $$\mathbf{C}^q = \{(\omega_1, \beta_1^q), \cdots, (\omega_M, \beta_M^q)\}, \text{ with rule weight } \theta^q, \tag{6}$$

where $x_1, x_2, \cdots, x_P$ represent the antecedent features and $\mathbf{A}^q = (A_1^q, A_2^q, \cdots, A_P^q)$ is the antecedent part of the belief rule R^q with each A_p^q belonging to fuzzy partitions $\{A_{p,1}, A_{p,2}, \cdots, A_{p,n_p}\}$ associated with p-th feature. β_m^q is the belief degree to which ω_m is believed to be the consequence for the q-th belief rule. In the belief structure, the consequence may be incomplete, i.e., $\sum_{m=1}^{M} \beta_m^q \leq 1$, and the left belief $1 - \sum_{m=1}^{M} \beta_m^q$ denotes the degree of global ignorance about the consequence. The rule weight θ^q with $0 \leq \theta^q \leq 1$, characterizes the certainty grade of the belief rule R^q.

Compared with the traditional fuzzy IF-THEN rule, the belief rule provides a more informative and realistic scheme for knowledge representation. By introducing belief function theory, the belief structure makes the rule more appropriate to characterize the uncertain information. For example, in generating each rule, only partial expert knowledge or limited training patterns are available. Thus, the corresponding consequence of this rule should not be complete. The belief rule structure can well characterize this type of incompleteness, with the left belief $1 - \sum_{m=1}^{M} \beta_m^q$ denoting the degree of global ignorance about the consequence induced by the partially available information.

III. Hybrid Belief Rule-Based Classification System

Considering the advantages of belief function theory for modeling and reasoning with uncertain information, in this section, a hybrid belief rule-based classification system (HBRBCS) is developed to integrate uncertain training data and expert knowledge to make classification. As shown Fig. 1, the proposed HBRBCS is composed of two components: a hybrid belief rule base (HBRB) and a belief reasoning & decision making module. The HBRB is constructed based on a data-driven belief rule base (DBRB) and a knowledge-driven belief rule base (KBRB). In our previous work [25], the problem of learning belief rule base from training data has been addressed. In Section III-A, we investigate how to learn from expert knowledge with the belief rule structure. In Section III-B, a fusion method is proposed to obtain an optimal HBRB by combining these two types of information via an optimization procedure. Sections III-C and III-D focus on how to make a reasoning based on the constructed HBRB and how to make a reasonable decision based on the reasoning results, respectively.

A. Knowledge-Driven Belief Rule Base (KBRB)

In knowledge-based system development, *knowledge representation* is the task of encoding expert knowledge into a knowledge base ready to use. Based on different types of applications, many knowledge representation schemes have been proposed, such as logical representations, production rules, semantic networks and structured frames [27]. For production rules, as knowledge is represented in the form of condition/action pairs, they provide a natural way to characterize the association between feature space and class space for classification problems. Thus, in this section, the production rules (especially, the fuzzy IF-THEN rules) are selected to represent expert knowledge.

Based on the fuzzy IF-THEN rules representation, expert knowledge is acquired from experts using the structured interview technique [28]. That is, experts are asked to assign fuzzy regions to each class and to give corresponding certainty grades. Accordingly, each piece of expert knowledge $\mathbf{e}_j$ can be represented as

$$\text{Expert Knowledge } \mathbf{e}_j: \quad \text{If } x_1 \text{ is } \mathbf{A}_1^j \text{ and } x_2 \text{ is } \mathbf{A}_2^j \text{ and } \cdots \text{ and } x_P \text{ is } \mathbf{A}_P^j, \text{ then the consequence is } \omega_j, \\ \text{with certainty grade } \theta_j, j = 1, 2, \cdots, M, \tag{7}$$

where $\mathbf{A}_p^j$ is subset of fuzzy partitions $\{A_{p,1}, A_{p,2}, \cdots, A_{p,n_p}\}$ associated with p-th feature, $p = 1, 2, \cdots, P$.

With the above M pieces of expert knowledge $\mathbf{e}_j$, $j = 1, 2, \cdots, M$, the problem now is how to generate a belief rule base with the belief rule structure as Eq. (6) from the expert knowledge. The knowledge-driven belief rule base is constructed in the following two stages: first, to expand the expert knowledge $\mathbf{e}_j$, $j = 1, 2, \cdots, M$ into belief rules by enumerating all the possible antecedent conditions; then, to combine those expanded rules having the same antecedent part.

1) Expansion of expert knowledge

For the belief rule as Eq. (6), each feature is associated with a single fuzzy partition, whereas for the expert knowledge in Eq. (7), each feature is associated with a set of fuzzy partitions. Thus, one piece of expert knowledge $\mathbf{e}_j$ can be expanded into some belief rules with the same consequent part and rule weight by enumerating all of

the possible antecedent conditions as follows.

$$
\begin{aligned}
\text{Belief Rule } R_j^1: \quad & \text{If } x_1 \text{ is } A_1^1 \text{ and } x_2 \text{ is } A_2^1 \text{ and} \cdots \text{ and } x_P \text{ is } A_P^1, \text{ then the consequence is} \\
& \mathbf{C}^1 = \{(\omega_j, 1)\}, \text{ with rule weight } \theta_j, \\
& \vdots \\
\text{Belief Rule } R_j^q: \quad & \text{If } x_1 \text{ is } A_1^q \text{ and } x_2 \text{ is } A_2^q \text{ and} \cdots \text{ and } x_P \text{ is } A_P^q, \text{ then the consequence is} \\
& \mathbf{C}^q = \{(\omega_j, 1)\}, \text{ with rule weight } \theta_j, \\
& \vdots \\
\text{Belief Rule } R_j^{Q_j}: \quad & \text{If } x_1 \text{ is } A_1^{Q_j} \text{ and } x_2 \text{ is } A_2^{Q_j} \text{ and} \cdots \text{ and } x_P \text{ is } A_P^{Q_j}, \text{ then the consequence is} \\
& \mathbf{C}^{Q_j} = \{(\omega_j, 1)\}, \text{ with rule weight } \theta_j,
\end{aligned}
\tag{8}
$$

where the antecedent parts $(A_1^1, A_2^1, \cdots, A_P^1), \cdots, (A_1^q, A_2^q, \cdots, A_P^q), \cdots, (A_1^{Q_j}, A_2^{Q_j}, \cdots, A_P^{Q_j})$ are all the possible combinations of different partitions for $\mathbf{A}_1^j, \mathbf{A}_2^j, \cdots, \mathbf{A}_P^j$, and Q_j is the number of belief rules generated from expert knowledge $\mathbf{e}_j$, with $Q_j = \prod_{p=1}^{P} |\mathbf{A}_p^j|$.

In the same way, all the M pieces of expert knowledge $\mathbf{e}_j$, $j = 1, 2, \cdots, M$ can be expanded to generate $\sum_{j=1}^{M} Q_j$ belief rules. However, as different classes may overlap in feature space, the belief rules generated from different pieces of expert knowledge may have the same antecedent part but different consequent parts. In other words, these rules are in conflict with each other. In the following, a combination method is provided to combine these conflicting rules by considering their rule weights.

2) Combination of conflicting rules

Suppose $R_{j_1}^{q_1}, \cdots, R_{j_{M'}}^{q_{M'}}$ $(2 \leq M' \leq M)$ are M' generated rules with the same antecedent part but different consequent parts $\{(\omega_{j_1}, 1)\}, \cdots, \{(\omega_{j_{M'}}, 1)\}$. In order to generate a compact KBRB, these M' rules should be fused into a new rule. The antecedent part of this new rule keeps the same, while its consequent part is obtained by combining those of the M' conflicting rules.

Each conflicting rule $R_{j_m}^{q_m}$ provides a piece of evidence that supports the class ω_{j_m} as the consequent part of the new fused rule. Considering that this rule has a certainty grade θ_{j_m}, this piece of evidence can be represented by a BBA m^{q_m} verifying:

$$
\begin{cases}
\mathrm{m}^{q_m}(\{\omega_{j_m}\}) &= \theta_{j_m} \\
\mathrm{m}^{q_m}(\Omega) &= 1 - \theta_{j_m} \\
\mathrm{m}^{q_m}(A) &= 0, \quad \forall A \in 2^{\Omega} \setminus \{\Omega, \{\omega_{j_m}\}\},
\end{cases}
\tag{9}
$$

where $\Omega = \{\omega_1, \cdots, \omega_M\}$ is the frame of discernment.

In a similar way, M' pieces of evidence $m^{q_1}, \cdots, m^{q_{M'}}$ can be constructed from the M' conflicting rules. These pieces of evidence are combined using Dempster's rule as follows.

$$
\begin{aligned}
\mathrm{m}(\{\omega_{j_m}\}) &= \tfrac{\theta_{j_m}}{1-K} \prod_{r \neq m} (1 - \theta_{j_r}), \quad m = 1, \cdots, M', \\
\mathrm{m}(\Omega) &= \tfrac{1}{1-K} \prod_{r=1}^{M'} (1 - \theta_{j_r}),
\end{aligned}
\tag{10}
$$

where K is the total conflicting belief mass

$$K = 1 - \prod_{r=1}^{M'} (1 - \theta_{j_r}) - \sum_{m=1}^{M'} \theta_{j_m} \prod_{r \neq m} (1 - \theta_{j_r}). \tag{11}$$

In addition, as the weights of the M' conflicting rules have already been considered in the combination process, the weight of the new fused rule is set to 1. In other words, in the combination process, the uncertainty of the M' conflicting rules is transferred to the consequence of the new fused rule. Therefore, the M' conflicting rules $R_{j_1}^{q_1}, \cdots, R_{j_{M'}}^{q_{M'}}$ can be replaced by a new fused rule with the same antecedent part and full weight but a different consequent part as $\{(\omega_{j_1}, \mathrm{m}(\{\omega_{j_1}\})), \cdots, (\omega_{j_{M'}}, \mathrm{m}(\{\omega_{j_{M'}}\}))\}$. In a similar way, all other sets of conflicting rules can be replaced by a series of new fused rules and then a compact KBRB is obtained to encode the expert knowledge about the classification problem.

B. Hybrid Belief Rule Base (HBRB)

In the previous section, a KBRB, dependent from the DBRB, is constructed based on expert knowledge. This section aims to fuse these two different belief rule bases into a new hybrid belief rule base for later reasoning process. In real-world problems, both training data and expert knowledge may be uncertain, i.e., these two types of information are not totally reliable. The uncertainty of training data comes from measurement noises or data entry errors. The uncertainty of expert knowledge is mainly due to limited or uncorrect assessment for the considered problem. Consequently, both the DBRB and KBRB only provide partially reliable information for the classification problem. Thus, in order to get a more powerful HBRB, we should take into account the weights of the DBRB and KBRB which reflect their different roles in the fusion process.

1) Fusion of DBRB and KBRB

Assuming the DBRB is β ($\beta > 0$) times as important as the KBRB, the weights of these two BRBs are set to $\beta/(1 + \beta)$ and $1/(1 + \beta)$, respectively. For notational convenience, we write $\lambda = \beta/(1 + \beta)$. Accordingly, the weight of the DBRB is λ and the weight of the KBRB is $1 - \lambda$ with $0 < \lambda < 1$. The adjustment factor λ plays an important role in adjusting the hybrid decision boundaries. With a large value of λ, the hybrid boundaries tend toward the DBRB boundaries. In contrast, with a small value of λ, the hybrid boundaries tend toward the KBRB boundaries.

With the above defined weights, we now fuse the Q_D data-driven belief rules (R_D^i, $i = 1, 2, \cdots, Q_D$) in the DBRB with the Q_K knowledge-driven belief rules (R_K^j, $j = 1, 2, \cdots, Q_K$) in the KBRB. As illustrated in Fig. 2, due to the partial information provided by both training data and expert knowledge, the generated rules in both the DBRB and KBRB only cover partial fuzzy regions. Furthermore, because of the independence between training data and expert knowledge, the fuzzy regions covered by the DBRB and the KBRB may not fully overlap. Thus, the rules in the integrated HBRB can be divided into the following three categories: rules with fuzzy regions only covered by the DBRB, rules with fuzzy regions only covered by the KBRB, and rules with fuzzy regions covered by both the DBRB and KBRB.

Let $\mathcal{S} = \{(i,j) \mid R_D^i \text{ and } R_K^j \text{ have the same antecedent fuzzy region}, i = 1, 2, \cdots, Q_D, j = 1, 2, \cdots, Q_K\}$, $\mathcal{S}_D = \{i \mid (i,j) \in \mathcal{S}, j = 1, 2, \cdots, Q_K\}$ and $\mathcal{S}_K = \{j \mid (i,j) \in \mathcal{S}, i = 1, 2, \cdots, Q_D\}$. The rules in the integrated HBRB are generated as follows.

- The rules with fuzzy regions only covered by the DBRB can be generated by assigning the corresponding rules in the DBRB with new weights $\lambda\theta^i$, $i \in \{1, 2, \cdots, Q_D\} \setminus \mathcal{S}_D$.
- The rules with fuzzy regions only covered by the KBRB can be generated by assigning the corresponding rules in the KBRB with new weights $(1 - \lambda)\theta^j$, $j \in \{1, 2, \cdots, Q_K\} \setminus \mathcal{S}_K$.
- The rules with fuzzy regions covered by both the DBRB and KBRB are generated by assigning the corresponding rules in both the DBRB and KBRB with new weights $\lambda\theta^i + (1 - \lambda)\theta^j$ and new consequences m^{ij} calculated by

$$\mathrm{m}^{ij} = {}^{\lambda}\mathrm{m}^i \oplus {}^{(1-\lambda)}\mathrm{m}^j, \quad (i,j) \in \mathcal{S}, \tag{12}$$

where ${}^{\lambda}\mathrm{m}^i$ denotes the discounted BBA for the consequence of the corresponding rule in DBRB with reliability factor λ, ${}^{(1-\lambda)}\mathrm{m}^j$ denotes the discounted BBA for the consequence of the corresponding rule in KBRB with reliability factor $(1 - \lambda)$, and $\oplus$ is Dempster's rule of combination.

Proposition 1: The generated HBRB reduces to DBRB and KBRB when the adjustment factor λ takes 1 and 0, respectively.

Proof: Suppose the adjustment factor $\lambda = 1$. First, for those rules with fuzzy regions only covered by the DBRB, the new assigned weights $\lambda\theta^i = \theta^i$, $i \in \{1, 2, \cdots, Q_D\} \setminus \mathcal{S}_D$. Thus, this category of rules are kept unchanged.

Second, for those rules with fuzzy regions only covered by the KBRB, the new assigned weights $(1 - \lambda)\theta^j = 0$, $j \in \{1, 2, \cdots, Q_K\} \setminus \mathcal{S}_K$. Thus, this category of rules is excluded from the generated HBRB.

Third, for those rules with fuzzy regions covered by both the DBRB and KBRB, the new assigned weights $\lambda\theta^i + (1 - \lambda)\theta^j = \theta^i$ and new consequences

$$\mathrm{m}^{ij} = {}^{\lambda}\mathrm{m}^i \oplus {}^{(1-\lambda)}\mathrm{m}^j = {}^{1}\mathrm{m}^i \oplus {}^{0}\mathrm{m}^j. \tag{13}$$

From the definition of Shafer's discounting operation in Eq. (4), it is easy to see that ${}^{1}\mathrm{m}^i = \mathrm{m}^i$ and, ${}^{0}\mathrm{m}^j$ becomes a vacuous BBA. Further, from the definition of Dempster's rule of combination in Eq. (5), the combination of any BBA with a vacuous BBA is equal to itself. Therefore, the new consequences $\mathrm{m}^{ij} = \mathrm{m}^i$, $(i,j) \in \mathcal{S}$. Thus, those rules in overlapping fuzzy regions inherit directly from the corresponding rules in the DBRB.

Consequently, it can be concluded that the generated HBRB reduces to DBRB when the adjustment factor $\lambda = 1$. In a similar way, it can be proved that the generated HBRB reduces to KBRB when the adjustment factor $\lambda = 0$.

∎

2) Optimization of the adjustment factor

In the above HBRB generation process, the rules from the DBRB and KBRB are combined to get an integrated HBRB that can make use of the information from both training data and expert knowledge. In this combination

process, the adjustment factor λ is used to adjust the weights of these two types of information. The adjustment factor λ can be specified by the user by evaluating the relative reliability of these two types of information. However, due to the ignorance about the quality of training data or expert knowledge, it may be difficult for the user to specify a proper value for λ. In the following, an optimization method for the adjustment factor is proposed by minimizing the average "leave-one-out" (LOO) test error.

Suppose that a labeled training set $T = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ is available. Let us consider a training vector $\mathbf{x}_i$ belonging to class ω_m. Take $\mathbf{x}_i$ as a test sample, and $T_i = T \setminus \{\mathbf{x}^i\}$ as the new training set. A HBR-B can be generated based on the new training set T_i and the expert knowledge. Using the belief reasoning and decision making method developed in the next section, one can get the LOO test output vector for $\mathbf{x}_i$ as $\mathbf{P}_i = (\text{BetP}_i(\{\omega_1\}), \cdots, \text{BetP}_i(\{\omega_M\}))$. Ideally, the classification output vector $\mathbf{P}_i$ should be as close as possible to the real class vector $\mathbf{t}_i = (t_{i1}, \cdots, t_{iM})$ (each binary indicator variable t_{ij} is defined by $t_{ij} = 1$, if $j = m$ and $t_{ij} = 0$, otherwise), with closeness being defined according to the squared error $E^\lambda(\mathbf{x}_i)$

$$E^\lambda(\mathbf{x}_i) = (\mathbf{P}_i - \mathbf{t}_i)(\mathbf{P}_i - \mathbf{t}_i)^{\mathrm{T}} = \sum_{j=1}^{M} (\text{BetP}_i(\{\omega_j\}) - t_{ij})^2. \tag{14}$$

The mean squared error over the whole training set T of size N is finally equal to

$$E^\lambda = \frac{1}{N} \sum_{i=1}^{N} E^\lambda(\mathbf{x}_i). \tag{15}$$

Therefore, the optimal value for λ is chosen with minimum LOO test error, i.e.,

$$\hat{\lambda} = \arg \min_{0 \le \lambda \le 1} E^\lambda. \tag{16}$$

As the minimization of E^λ is performed with respect to a single parameter in a bounded domain, a very simple search procedure can be used.

C. Reasoning in Belief Function Framework

Traditionally, for rule-based systems, the single-winner rule [29] is usually used to make inference. However, when both training data and expert knowledge have low quality, the generated rules in HBRB have high unreliability. Under this circumstance, single-winner reasoning method may entail great risk. In order to get a more robust classification result, in this section, a belief reasoning method is used to combine the consequences of all of the rules activated by the query pattern.

1) Association degree calculation

In reasoning process, the generated rules in HBRB play different roles in determining the class of the input query pattern. This type of effect is characterized by the association degree between the query pattern and the consequence of each rule.

Suppose $\mathbf{y} = (y_1, y_2, \cdots, y_P)$ is a query pattern to be classified. The association degree between query pattern $\mathbf{y}$ and the consequence of belief rule R^q is defined as

$$\alpha^q = \left[\prod_{p=1}^{P} \mu_{A_p^q}(y_p) \right]^{1/P} \cdot \theta^q. \tag{17}$$

where $\mu_{A_p^q}$ is the membership function of fuzzy set A_p^q. It can be seen that the association degree is determined by two factors: the matching degree $\left[\prod_{p=1}^{P} \mu_{A_p^q}(y_p)\right]^{1/P}$ and the rule weight θ^q. The matching degree reflects the similarity between the query pattern and the antecedent part of the belief rule, whereas the rule weight characterizes the reliability of the belief rule itself.

2) Belief reasoning algorithm

In HBRB, only those activated rules that have nonzero association degrees with query pattern **y** are useful for classification. The association degrees reflect the strength of activated rules in determining the class of query pattern **y**. In belief function framework, the association degree α of each activated rule can be taken into account using Shafer's discounting operation as

$$
\begin{aligned}
{}^{\alpha}\mathrm{m}(\{\omega_m\}) &= \alpha \cdot \beta_m, \quad m = 1, 2, \cdots, M, \\
{}^{\alpha}\mathrm{m}(\Omega) &= \alpha \cdot \beta_\Omega + (1 - \alpha)
\end{aligned}
\tag{18}
$$

Suppose that there are L belief rules being activated by query pattern **y**. With the above formula, L corresponding discounted BBAs ${}^{\alpha}\mathrm{m}_i$, $i = 1, 2, \cdots, L$, can be obtained.

In order to make a decision regarding the discounted consequences of activated belief rules, the corresponding BBAs can be combined using Dempster's rule. Considering the fact that the focal elements of each associated BBA are all singletons except the ignorance set Ω, the computational complexity of Dempster's rule defined in Eq. (5) is reduced to linear time. As Dempster's rule is both commutative and associative, these L BBAs can be combined recursively. Define $I(i)$ as the index set of the former i BBAs. Let $\mathrm{m}_{I(i)}$ be the BBA after combining all of the former i BBAs. Then $\mathrm{m}_{I(i+1)}$ is obtained by combining $\mathrm{m}_{I(i)}$ with $(i+1)$th discounted BBA ${}^{\alpha}\mathrm{m}_{i+1}$ as follows

$$
\begin{aligned}
\mathrm{m}_{I(i+1)}(\{\omega_m\}) &= K_{I(i+1)} \left[\mathrm{m}_{I(i)}(\{\omega_m\}) \cdot {}^{\alpha}\mathrm{m}_{i+1}(\{\omega_m\}) + \mathrm{m}_{I(i)}(\Omega) \cdot {}^{\alpha}\mathrm{m}_{i+1}(\{\omega_m\}) \right. \\
&\quad \left. + \mathrm{m}_{I(i)}(\{\omega_m\}) \cdot {}^{\alpha}\mathrm{m}_{i+1}(\Omega) \right], \quad m = 1, 2, \cdots, M \\
\mathrm{m}_{I(i+1)}(\Omega) &= K_{I(i+1)} \left[\mathrm{m}_{I(i)}(\Omega) \cdot {}^{\alpha}\mathrm{m}_{i+1}(\Omega) \right] \\
& i = 1, 2, \cdots, L - 1,
\end{aligned}
\tag{19}
$$

where $K_{I(i+1)}$ is a normalizing factor defined as

$$
K_{I(i+1)} = \left[1 - \sum_{j=1}^{M} \sum_{p=1, p \neq j}^{M} \mathrm{m}_{I(i)}(\{\omega_j\}) \cdot {}^{\alpha}\mathrm{m}_{i+1}(\{\omega_p\}) \right]^{-1}.
\tag{20}
$$

Note that $\mathrm{m}_{I(1)}(\{\omega_m\}) = {}^{\alpha}\mathrm{m}_1(\{\omega_m\})$ for $m = 1, 2, \cdots, M$, and $\mathrm{m}_{I(1)}(\Omega) = {}^{\alpha}\mathrm{m}_1(\Omega)$. This recursive combination can initiate with the first BBA. Accordingly, when the recursive index i reaches $L - 1$, the final results $\mathrm{m}_{I(L)}(\{\omega_m\})$, $m = 1, 2, \cdots, M$, and $\mathrm{m}_{I(L)}(\Omega)$ are obtained by combining all of the L BBAs. This combination result is the basis for the later decision process.

D. Decision Making Strategies

For decision making based on the reasoning result calculated by Eq. (19), the corresponding belief function Bel, plausibility function Pl and pignistic probability BetP are useful standards of measurement. The credibility,

plausibility and pignistic probability of each class ω_m are calculated as follows

$$
\begin{aligned}
\text{Bel}(\{\omega_m\}) &= \text{m}_{I(L)}(\{\omega_m\}) \\
\text{Pl}(\{\omega_m\}) &= \text{m}_{I(L)}(\{\omega_m\}) + \text{m}_{I(L)}(\Omega) \\
\text{BetP}(\{\omega_m\}) &= \text{m}_{I(L)}(\{\omega_m\}) + \frac{\text{m}_{I(L)}(\Omega)}{M} \\
& \quad m = 1, 2, \cdots, M.
\end{aligned}
\tag{21}
$$

For different classification problems, different types of decisions may be required based on the user's requirement. For most cases, the user wishes to get a single class output for one query pattern $\mathbf{y}$. However, for some specific applications (e.g., target recognition [9]), in contrast to hard decision that assigns a single class to a pattern, the user sometimes may prefer a soft decision providing multiple decision options for further analysis, which is believed to be better than producing a wrong classification. The belief function theory provides power tools to develop both the hard and soft decision strategies.

1) Hard decision strategy

For decision making with hard partition, the pignistic probability BetP which transforms the belief distribution into a probability distribution, is a common alternative. Based on the pignistic probability BetP, the single class output for one query pattern is decided as

$$
\omega = \arg \max_{\omega_m \in \Omega} \text{BetP}(\{\omega_m\}).
\tag{22}
$$

That is, the query pattern is assigned to the class with the maximum pignistic probability.

2) Soft decision strategy

As introduced in Section II-A, the belief function Bel and plausibility function Pl provide the lower and upper probabilities, respectively. That is, the probability distribution P over Ω satisfying:

$$
\text{Bel}(\{\omega_m\}) \leq \text{P}(\{\omega_m\}) \leq \text{Pl}(\{\omega_m\}), \quad \text{for } \omega_m \in \Omega.
\tag{23}
$$

Fig. 3 shows the credibility, plausibility, and probability for all of the M classes in the real interval $[0, 1]$, where $\omega^{(1)}$ and $\omega^{(M)}$ are the classes having the maximum and minimum credibility or plausibility, respectively (As seen from Equation (21), the credibility and plausibility synchronously get the maximum and minimum values).

The basic idea of the soft decision strategy is that if the probabilities of two classes overlap with each other, these two classes are undistinguishable. Thus, the result of the soft decision is a set containing all of the classes that are undistinguishable with the topmost class $\omega^{(1)}$. Table I shows the pseudo-code of the soft decision strategy.

To better illustrate how the hard and soft decision strategies work, the following example is considered.

Example 1. Let's consider the frame of classes $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. For a query pattern $\mathbf{y}$, it is supposed that the reasoning result using Eq. (19) is

$$
\text{m}: \quad \text{m}(\{\omega_1\}) = 0, \quad \text{m}(\{\omega_2\}) = 0.4, \quad \text{m}(\{\omega_3\}) = 0.3, \quad \text{m}(\{\omega_4\}) = 0.1, \quad \text{m}(\Omega) = 0.2.
$$

Based on the above BBA, the corresponding credibility, plausibility, and pignistic probability of each class are calculated as

TABLE I

THE SOFT DECISION STRATEGY

Preprocess: Sort the M classes in a descending order of credibility or plausibility.

Initialization: Set the initial decision set $W = \{\omega^{(1)}\}$.

For $i = 2$ to M

    If $\quad \mathrm{Pl}(\{\omega^{(i)}\}) > \mathrm{Bel}(\{\omega^{(1)}\})$

        $W = W \cup \{\omega^{(i)}\}$;

    Else

        Exit.

End

	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$	$\{\omega_4\}$
Bel	0	0.4	0.3	0.1
Pl	0.2	0.6	0.5	0.3
BetP	0.05	0.45	0.35	0.15

With hard decision strategy, based on Eq. (23), it is easy to get the single class output ω_2, which has the maximum pignistic probability.

With soft decision strategy, first the four classes are sorted in a descending order of credibility or plausibility as

$$\omega_2 \succ \omega_3 \succ \omega_4 \succ \omega_1.$$

According to the soft decision strategy in Table I, the initial decision set only contains the topmost class ω_2. As $\mathrm{Pl}(\{\omega_3\}) > \mathrm{Bel}(\{\omega_2\})$, the class ω_3 should also be included in the decision set. But because $\mathrm{Pl}(\{\omega_4\}) < \mathrm{Bel}(\{\omega_2\})$, the decision set should not contain the class ω_4 and there is no necessity to further consider the later class ω_1. Finally, the soft decision result is $\{\omega_2, \omega_3\}$.

IV. NUMERICAL STUDY

In order to present the implementation of the proposed HBRBCS and demonstrate its capacity of combining both uncertain training data and expert knowledge to make classification, a numerical study for an airborne target classification in the air surveillance system is provided.

A. Problem Description

For air surveillance systems [30], one of the most important tasks is to correctly recognize noncooperative flying objects within their surveillance volume. In general, target classification is based on a set of features or attributes which distinguish targets according to their shapes or kinematic behaviors. To fully exploit the feature space, a surveillance system often consists of multiple sensors. For example, a radar can provide kinematic features (e.g., speed, acceleration) and an infrared sensor can supply shape features such as the length. In this study, we consider the

TABLE II

FEATURE INTERVALS FOR THREE AIRBORNE TARGET CLASSES

Class	AveSpeed (km/h)	MaxAcc (g)	AveLength (m)
Commercial (ω_1)	[600, 800]	[0, 1]	[25, 65]
Bomber (ω_2)	[400, 700]	[0, 4]	[15, 45]
Fighter (ω_3)	[500, 1000]	[0, 6]	[10, 30]

classification of targets in predefined categories $\{Commercial\ plane, Bomber, Fighter\}$, based on their average speed (*AveSpeed*), maximum acceleration (*MaxAcc*) and average length (*AveLength*) measured by a multi-sensor system composed of a land-based radar and an airborne infrared sensor.

In this numerical study, the feature measurements are simulated using the Gaussian-distributed class-conditioned probability functions. The Gaussian densities are used, with the parameters selected in such a way that $P\{s_{\min} < x < s_{\max}\} = 0.95$, where $[s_{\min}, s_{\max}]$ is the feature interval given in Table II. Fig. 4 shows the distributions of the three features conditioned on the class.

B. Implementation of the Hybrid Belief Rule Base

In this study, assume that 120 labeled training patterns $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})$, $i = 1, 2, \cdots, 120$, are collected by the multi-sensor system according to the above class-conditioned probability distributions with equal prior probabilities (1/3, 1/3, 1/3). Further, assume the available training patterns are not fully reliable, i.e., some of them have wrong class labels. This scenario is simulated by adding class noise with noise level of $x\%$ indicating that $x\%$ of the samples in the training set are mislabeled. The class labels of these samples are randomly changed to different ones within the domain of the class. Apart from the uncertain training data, suppose that three pieces of partially reliable expert knowledge $\mathbf{e}_j$, $j = 1, 2, \cdots, 3$, are obtained with the structured interview. These two types of information are used to construct the HBRB using the proposed method. The processes of constructing the HBRB concerning the airborne target classification problem are implemented as follows.

Step 1: Preprocessing: fuzzification of the feature space

The prerequisite step to generate a belief rule base is to fuzzify the feature space. The fuzzy grid partition method [31] is used to fuzzify the feature space. Suppose according to the *a priori* knowledge, it is known that the three features *AveSpeed*, *MaxAcc*, and *AveLength*, change in the intervals $[400, 1000]$, $[0, 6]$ and $[10, 70]$, respectively. The partition number for each feature is set to three. As only a few training patterns are available, a large partition number may result in over-fitting. Moreover, a relatively small partition number also makes it easier for the experts to assign fuzzy regions to each class. Based on the fuzzy grid partition method, the fuzzification of the three features is shown in Fig. 5.

Step 2: Construction of DBRB

In this step, 120 labeled training patterns $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})$, $i = 1, 2, \cdots, 120$, are used to construct the DBRB using the method developed in [25]. Table III shows the DBRB containing 12 belief rules learnt from the uncertain

TABLE III

DBRB CONSTRUCTED BASED ON THE UNCERTAIN TRAINING DATA

Rule number	Rule weight	Antecedent	Consequent
1	0.45	$L \wedge L \wedge L$	$\{(\omega_1, 0), (\omega_2, 0.3257), (\omega_3, 0.4812)\}$
2	0.69	$L \wedge L \wedge M$	$\{(\omega_1, 0.0041), (\omega_2, 0.8833), (\omega_3, 0.1126)\}$
3	0.36	$L \wedge M \wedge L$	$\{(\omega_1, 0), (\omega_2, 0), (\omega_3, 0.3581)\}$
4	0.56	$L \wedge M \wedge M$	$\{(\omega_1, 0), (\omega_2, 0.6687), (\omega_3, 0)\}$
5	0.90	$M \wedge L \wedge L$	$\{(\omega_1, 0), (\omega_2, 0.1011), (\omega_3, 0.8879)\}$
6	1.00	$M \wedge L \wedge M$	$\{(\omega_1, 0.8632), (\omega_2, 0.1245), (\omega_3, 0)\}$
7	0.54	$M \wedge L \wedge H$	$\{(\omega_1, 0.7906), (\omega_2, 0.1047), (\omega_3, 0.1035)\}$
8	0.44	$M \wedge M \wedge L$	$\{(\omega_1, 0), (\omega_2, 0.7084), (\omega_3, 0.2119)\}$
9	0.69	$M \wedge M \wedge M$	$\{(\omega_1, 0), (\omega_2, 0.8998), (\omega_3, 0)\}$
10	0.63	$H \wedge L \wedge L$	$\{(\omega_1, 0.2007), (\omega_2, 0), (\omega_3, 0.7591)\}$
11	0.43	$H \wedge M \wedge L$	$\{(\omega_1, 0.1171), (\omega_2, 0), (\omega_3, 0.8295)\}$
12	0.36	$H \wedge M \wedge M$	$\{(\omega_1, 0), (\omega_2, 0), (\omega_3, 0.6657)\}$

training data with class noise level of 30%. Although the DBRB generation method can reduce the adverse effects of class noises, the consequence of one rule may still be unreliable in small training set and excessively noisy conditions. For example, in the constructed DBRB, the rule R_D^3 assigns most belief to ω_3, which is not consistent with the real class-conditioned distributions shown in Fig. 4 (from which, class ω_2 should be assigned more belief). This is because only one training sample is assigned to the corresponding fuzzy region $\{L \wedge M \wedge L\}$, and this only training sample is not representative of the real class-conditioned distributions. Fortunately, the developed rule weight generation method only assigns this rule a small weight, which decreases its effect in the reasoning process.

Step 3: Construction of KBRB

Suppose the following three pieces of expert knowledge $\mathbf{e}_j$, $j = 1, 2, \cdots, 3$, are obtained by asking the experts to assign the fuzzy regions for each specific target class and to give the corresponding certainty grade.

$\mathbf{e}_1$: If x_1 is $\{M\}$ and x_2 is $\{L\}$ and x_3 is $\{M, H\}$, then consequence is ω_1,
　　　　with certainty grade 0.9,

$\mathbf{e}_2$: If x_1 is $\{L, M\}$ and x_2 is $\{L, M\}$ and x_3 is $\{L, M\}$, then consequence is ω_2,
　　　　with certainty grade 0.7,

$\mathbf{e}_3$: If x_1 is $\{M, H\}$ and x_2 is $\{L, M, H\}$ and x_3 is $\{L\}$, then consequence is ω_3,
　　　　with certainty grade 0.8.

As indicated above, each piece of expert knowledge covers several fuzzy regions as shown in Table IV.

From Table IV, it can be seen that one fuzzy region $\{M \wedge L \wedge M\}$ is covered by both expert knowledge $\mathbf{e}_1$ and $\mathbf{e}_2$, and two fuzzy regions $\{M \wedge L \wedge L\}$ and $\{M \wedge M \wedge L\}$ are covered by both expert knowledge $\mathbf{e}_2$ and $\mathbf{e}_3$. For these three regions, the consequences are obtained by combing the conflicting pieces of expert knowledge considering their certainty grades with Eqs.(9-11). For those non-overlapping regions, the consequences are kept unchanged.

TABLE IV

FUZZY REGIONS COVERED BY EACH PIECE OF EXPERT KNOWLEDGE

Expert knowledge	Covered fuzzy regions
$\mathbf{e}_1$	$\overline{\{M \wedge L \wedge M\}}$, $\{M \wedge L \wedge H\}$
$\mathbf{e}_2$	$\{L \wedge L \wedge L\}$, $\{L \wedge M \wedge L\}$, $\{L \wedge L \wedge M\}$, $\{L \wedge M \wedge M\}$,
	$\underline{\{M \wedge L \wedge L\}}$, $\underline{\{M \wedge M \wedge L\}}$, $\overline{\{M \wedge L \wedge M\}}$, $\{M \wedge M \wedge M\}$
$\mathbf{e}_3$	$\underline{\{M \wedge L \wedge L\}}$, $\underline{\{M \wedge M \wedge L\}}$, $\{M \wedge H \wedge L\}$, $\{H \wedge L \wedge L\}$, $\{H \wedge M \wedge L\}$, $\{H \wedge H \wedge L\}$

TABLE V

KBRB CONSTRUCTED BASED ON THE EXPERT KNOWLEDGE

Rule number	Rule weight	Antecedent	Consequent
1	0.70	$L \wedge L \wedge L$	$\{(\omega_1, 0), (\omega_2, 1), (\omega_3, 0)\}$
2	0.70	$L \wedge L \wedge M$	$\{(\omega_1, 0), (\omega_2, 1), (\omega_3, 0)\}$
3	0.70	$L \wedge M \wedge L$	$\{(\omega_1, 0), (\omega_2, 1), (\omega_3, 0)\}$
4	0.70	$L \wedge M \wedge M$	$\{(\omega_1, 0), (\omega_2, 1), (\omega_3, 0)\}$
5	1.00	$M \wedge L \wedge L$	$\{(\omega_1, 0), (\omega_2, 0.3182), (\omega_3, 0.5455)\}$
6	1.00	$M \wedge L \wedge M$	$\{(\omega_1, 0.7297), (\omega_2, 0.1892), (\omega_3, 0)\}$
7	0.90	$M \wedge L \wedge H$	$\{(\omega_1, 1), (\omega_2, 0), (\omega_3, 0)\}$
8	1.00	$M \wedge M \wedge L$	$\{(\omega_1, 0), (\omega_2, 0.3182), (\omega_3, 0.5455)\}$
9	0.70	$M \wedge M \wedge M$	$\{(\omega_1, 0), (\omega_2, 1), (\omega_3, 0)\}$
10	0.80	$M \wedge H \wedge L$	$\{(\omega_1, 0), (\omega_2, 0), (\omega_3, 1)\}$
11	0.80	$H \wedge L \wedge L$	$\{(\omega_1, 0), (\omega_2, 0), (\omega_3, 1)\}$
12	0.80	$H \wedge M \wedge L$	$\{(\omega_1, 0), (\omega_2, 0), (\omega_3, 1)\}$
13	0.80	$H \wedge H \wedge L$	$\{(\omega_1, 0), (\omega_2, 0), (\omega_3, 1)\}$

Table V shows the KBRB containing 13 belief rules constructed from the expert knowledge. In the constructed KBRB, rule R_K^6 is obtained by combining the conflicting items $\mathbf{e}_1$ and $\mathbf{e}_2$, and rules R_K^5 and R_K^8 are obtained by combining the conflicting items $\mathbf{e}_2$ and $\mathbf{e}_3$. It can be seen that for the three new fused rules, the consequences are not complete (i.e., the sum of the belief for all of the three classes is less than one), with the left belief characterizing the global ignorance induced by the partially reliable expert knowledge. Due to partially available expert knowledge, the generated KBRB only covers part of the fuzzy regions of the feature space. In addition, due to the insufficiency of the expert knowledge, the consequence of one rule may be unreliable. For example, in the constructed KBRB, the rule R_K^1 assigns full belief to ω_2. However, according to the real class-conditioned distributions shown in Fig. 4, the samples from class ω_3 also have a high possibility to fall into the corresponding fuzzy region $L \wedge L \wedge L$.

Step 4: Construction of HBRB

In this stage, the rules from the DBRB (Table III) and KBRB (Table V) are combined based on the fusion algorithm developed in Section III-B. Table VI shows the optimal HBRB containing 14 belief rules. Compared to the previous generated DBRB and KBRB, the integrated HBRB has the following two main advantages:

TABLE VI

HBRB CONSTRUCTED BASED ON THE UNCERTAIN TRAINING DATA AND EXPERT KNOWLEDGE

Rule number	Rule weight	Antecedent	Consequent
1 $(R_D^1, R_K^1)^a$	0.52	$L \wedge L \wedge L$	$\{(\omega_1, 0), (\omega_2, 0.3853), (\omega_3, 0.2833)\}$
2 (R_D^2, R_K^2)	0.69	$L \wedge L \wedge M$	$\{(\omega_1, 0.0022), (\omega_2, 0.7346), (\omega_3, 0.0614)\}$
3 (R_D^3, R_K^3)	0.45	$L \wedge M \wedge L$	$\{(\omega_1, 0), (\omega_2, 0.2146), (\omega_3, 0.2053)\}$
4 (R_D^4, R_K^4)	0.60	$L \wedge M \wedge M$	$\{(\omega_1, 0), (\omega_2, 0.6263), (\omega_3, 0)\}$
5 (R_D^5, R_K^5)	0.93	$M \wedge L \wedge L$	$\{(\omega_1, 0), (\omega_2, 0.0930), (\omega_3, 0.6786)\}$
6 (R_D^6, R_K^6)	1.00	$M \wedge L \wedge M$	$\{(\omega_1, 0.6873), (\omega_2, 0.0918), (\omega_3, 0)\}$
7 (R_D^7, R_K^7)	0.63	$M \wedge L \wedge H$	$\{(\omega_1, 0.7976), (\omega_2, 0.0025), (\omega_3, 0.0018)\}$
8 (R_D^8, R_K^8)	0.59	$M \wedge M \wedge L$	$\{(\omega_1, 0), (\omega_2, 0.5153), (\omega_3, 0.2084)\}$
9 (R_D^9, R_K^9)	0.69	$M \wedge M \wedge M$	$\{(\omega_1, 0), (\omega_2, 0.8028), (\omega_3, 0)\}$
10 $(--, R_K^{10})$	0.22	$M \wedge H \wedge L$	$\{(\omega_1, 0), (\omega_2, 0), (\omega_3, 1)\}$
11 (R_D^{10}, R_K^{11})	0.68	$H \wedge L \wedge L$	$\{(\omega_1, 0.0004), (\omega_2, 0), (\omega_3, 0.8024)\}$
12 (R_D^{11}, R_K^{12})	0.53	$H \wedge M \wedge L$	$\{(\omega_1, 0.6039), (\omega_2, 0), (\omega_3, 0.7052)\}$
13 $(R_D^{12}, --)$	0.26	$H \wedge M \wedge M$	$\{(\omega_1, 0), (\omega_2, 0), (\omega_3, 0.6657)\}$
14 $(--, R_K^{13})$	0.22	$H \wedge H \wedge L$	$\{(\omega_1, 0), (\omega_2, 0), (\omega_3, 1)\}$

^{a}The corresponding rules in DBRB and KBRB with the same antecedent parts are shown in brackets.

1) It covers more fuzzy regions than either the DBRB or HBRB and in this case, it is more powerful to classify those patterns uncovered by either the DBRB or HBRB.

2) In the overlapping fuzzy regions of the DBRB and HBRB, through combination, the rules in HBRB reduced the potential unreliability existed in the corresponding rules of DBRB or KBRB. For example, as indicated in *Step 2*, the rule R_D^3 generated from the uncertain training data is unreliable, but after combination with the corresponding rule R_K^3, the consequence of the combined rule R_H^3 has better representation for the real class distributions in the fuzzy region $\{L \wedge M \wedge L\}$. Similarly, as indicated in *Step 3*, the rule R_K^1 generated from the uncertain expert knowledge is unreliable, but after combination with the corresponding rule R_D^1, a better rule R_H^1 is generated for the fuzzy region $\{L \wedge L \wedge L\}$.

C. Comparative Study

In this comparative study, the HBRBCS is compared to the DBRBCS ($\lambda = 1$, which only considers the uncertain training data) and the KBRBCS ($\lambda = 0$, which only considers the uncertain expert knowledge) under different noise levels for the training data. A test set of 3000 reliable samples drawn from the original class-conditioned probability distributions are used for error estimation. For the HBRBCS, the optimal adjustment factor λ by optimizing the average LOO test error is used to get the integrated HBRB, and in order to make a fair comparison, the hard decision strategy is selected. In addition, two robust data-based classifiers (EEkNN [32], BagC4.5[33]) and a representative hybrid classifier based on both training data and expert knowledge (AFRBCS [13]) are also considered in the comparison.

1) *EEkNN*: This method is based on evidential editing technique for the training data set. The edited training data set is then used to classify a query pattern based on k-nearest neighbor rule. In this experiment, only the results for the optimal value of k ($1 \leq k \leq 20$) observed in each case are reported.

2) *BagC4.5*: This is a multiple classifier system that considers decision tree learning method C4.5 as the base classifier. In this method, the bagging technique is used to resample the original training set, and then the base classifier is trained with different data sets. In this experiment, the replicate number is set to 10, and the majority voting rule is used.

3) *AFRBCS*: This is a hybrid classifier based on both training data and expert knowledge. In this method, a fuzzy rule-based classification system is first built based on expert knowledge, and then the parameters of the used fuzzy membership function are optimized based on training data. In this experiment, the conservative parameter is set to 0.5.

Table VII shows the classification error rates for different methods with different noise levels. With the increase of the class noise in the training data set, the performance of all of the three data-based classifiers EEkNN, BagC4.5 and DBRBCS, decrease, whereas the DBRBCS shows more robust to class noise due to the utilized belief rule structure and belief reasoning method. The KBRBCS, which classifies query patterns only based on the expert knowledge, always yields a moderate performance. Interestingly, the HBRBCS outperforms both the DBRBCS and KBRBCS with any noise level. The reason is that, on the one hand, the fused HBRB covers more fuzzy regions than the partial DBRB and KBRB and, on the other hand, for the overlapping fuzzy regions, thanks to the proposed fusion procedure, the rules in HBRB reduced the potential unreliability existed in the corresponding rules of DBRB or KBRB. Besides, by comparing the two hybrid classifiers AFRBCS and HBRBCS, it can be seen that the proposed HBRBCS yields better performance, especially for cases with high data noise levels. It is because in HBRBCS the weights of training data and expert knowledge are adjusted adaptively according to the qualities of these two types of information. Whereas, the AFRBCS always uses training data to update the knowledge-based model, no matter the available training data set is reliable or not.

In order to find out whether significant differences exist among different methods, error estimation confidence intervals (with confidence level $A = 95\%$) [1] for the best method are shown in the last column. It is interesting to note that only when the noise level is quite low (0% and 10%), the error rate of the second best method is within the corresponding confidence interval. When the class noise increases, the best HBRBCS method shows a statistically significant advantage. Therefore, the classification performance can improve greatly by making use of the complementary information from uncertain training data and expert knowledge based on belief function theory, especially when both sources of information have high uncertainty.

---

[1]Computed by numerically solving the equations: $\sum_{k \geq K} P(k, n, \underline{p}) = (1-A)/2$ and $\sum_{k \leq K} P(k, n, \overline{p}) = (1-A)/2$, where $P(k, n, p)$ is the binomial distribution, n is the number of test patterns, K is the number of patterns misclassified, A is the confidence level, and $[\underline{p}, \overline{p}]$ is the confidence interval.

TABLE VII

CLASSIFICATION ERROR RATES (IN %) FOR CONSIDERED METHODS WITH DIFFERENT NOISE LEVELS

Noise Level	EEkNN	BagC4.5	DBRBCS	KBRBCS	AFRBCS	HBRBCS	CI[a]
0%	17.40	18.63	17.87	23.37	18.23	**17.30**[b]	[15.97,18.71]
10%	21.27	21.20	18.83	23.37	20.10	**17.83**	[16.48,19.24]
20%	24.10	22.07	20.47	23.37	22.20	**18.47**	[17.10,19.92]
30%	29.33	26.33	24.03	23.37	23.93	**18.83**	[17.44,20.28]
40%	37.23	33.70	30.70	23.37	26.67	**19.47**	[18.07,20.92]
50%	41.40	40.13	38.13	23.37	30.47	**20.40**	[18.97,21.91]

[a]The last column is the 95% confidence interval of the best method.

[b]Results in boldface correspond to the lowest error rate.

TABLE VIII

COMPARISON OF THE ESTIMATED AND TESTED OPTIMAL λ AS WELL AS THEIR CORRESPONDING CLASSIFICATION ERROR RATES (IN %) WITH DIFFERENT NOISE LEVELS

Noise Level	0%	10%	20%	30%	40%	50%
Est. λ	0.80	0.77	0.78	0.73	0.68	0.56
Tes. λ	0.82	0.79	0.75	0.71	0.64	0.51
Error rate with est. λ	17.30	17.83	18.47	18.83	19.47	20.40
Error rate with tes. λ	17.27	17.80	18.40	18.73	19.16	19.69

D. Parameter Analysis

The adjustment factor λ plays an important role in determining the classification performance of the HBRBCS. This section gives an analysis for the effect of the adjustment factor λ and evaluates whether the optimization method developed in Section III-B works well.

Fig. 6 shows the classification error rate of the HBRBCS with the adjustment factor ranging from 0 to 1 under different noise levels. It can be seen that the optimal values of the adjustment factor λ are different under different noise levels. With the increase of the noise level, the optimal value of λ tends to be smaller. This is because, in these cases, the DBRB generated from the noisy training data becomes less reliable and the KBRB which is not affected by the noisy training data takes more important role in determining the classification.

Table VIII compares the estimated optimal value of λ by optimizing the average LOO test error with the tested optimal value of λ as well as their corresponding classification error rates. It can be seen that the parameter optimization method provides a good estimation for the optimal value of λ. Accordingly, the classification error rate with the estimated optimal value of λ is very close to the tested optimal error rate. In addition, the estimation is more accurate with low noise levels, because in these cases the training data has a good representation of the real class-conditioned distributions.

V. Conclusion

In order to make use of the information from both uncertain training data and expert knowledge for classification, a hybrid belief rule-based classification system (HBRBCS) has been developed based on the belief rule structure. The proposed HBRBCS offers complementary advantages from data-driven models and knowledge-driven models. This system can be useful for many real-world applications where both uncertain training data and expert knowledge are available. An optimal hybrid belief rule base (HBRB) was learnt from uncertain training data and expert knowledge to establish the association between the feature space and the class space. Based on the generated HBRB, a belief reasoning algorithm was applied to classify a query pattern in a robust way and both a hard and soft decision strategies were provided for user to select depending on the different needs. An airborne target classification in the air surveillance system has been studied to present the implementation of the proposed HBRBCS and to demonstrate its capacity of combining both uncertain training data and expert knowledge to make classification. The experiment results showed that the HBRBCS can make good use of these two types of independent and complementary information and achieve better performance.

In this work, we focus on classification problems for static data sets. The uncertainty is mainly due to the impurity in data sets. In the experiments, we simulated the uncertainty by adding noise to the originally collected data. In contrast, for the classification of dynamic data streams, the behavior of data changing over time (abrupt change or gradual change) can also cause uncertainty. Several methods [34], [35], [36] have been developed to cope with the classification problems for uncertain data streams, whereas none of them made use of the available expert knowledge in the classification process. Therefore, a future research direction is to extend the proposed hybrid system to make classification based on uncertain data streams and expert knowledge.

References

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NJ: John Wiley & Sons, 2012.

[2] D. Dubois, P. Hájek, and H. Prade, "Knowledge-driven versus data-driven logics," *J. Logic Lang. Inform.*, vol. 9, no. 1, pp. 65–89, 2000.

[3] D. Codetta-Raiteri and L. Portinale, "Dynamic bayesian networks for fault detection, identification, and recovery in autonomous spacecraft," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 45, no. 1, pp. 13–24, 2015.

[4] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, MA: MIT Press, 2014.

[5] J. Xu, "Rule-based automatic software performance diagnosis and improvement," *Perform. Evaluation*, vol. 69, no. 11, pp. 525–550, 2012.

[6] C. S. Lee and M. H. Wang, "A fuzzy expert system for diabetes decision support application," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 1, pp. 139–153, 2011.

[7] D. D. Wu, D. L. Olson, and C. Luo, "A decision support approach for accounts receivable risk management," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 44, no. 12, pp. 1624–1632, 2014.

[8] P. Dasgupta, "A multiagent swarming system for distributed automatic target recognition using unmanned aerial vehicles," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 38, no. 3, pp. 549–563, 2008.

[9] L. Jiao, Q. Pan, X. Feng, and F. Yang, "A belief-rule-based inference method for carrier battle group recognition," in *Lecture Notes in Electrical Engineering*, Z. Sun and Z. Deng, Eds. Heidelberg, Germany: Springer, 2013, vol. 254, pp. 261–271.

[10] P. H. Tsai, Y. J. Lin, Y. Z. Ou, E. T. Chu, and J. W. Liu, "A framework for fusion of human sensor and physical sensor data," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 44, no. 9, pp. 1248–1261, 2014.

[11] Z. J. Zhou, C. H. Hu, J. B. Yang, D. L. Xu, M. Y. Chen, and D. H. Zhou, "A sequential learning algorithm for online constructing belief-rule-based systems," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1790–1799, 2010.

[12] W. Tang, K. Z. Mao, L. O. Mak, G. W. Ng, Z. Sun, J. H. Ang, and G. Lim, "Target classification using knowledge-based probabilistic model," in *Proc. 2011 FUSION*, Chicago, USA, 2011, pp. 1–8.

[13] W. Tang, K. Z. Mao, L. O. Mak, and G. W. Ng, "Adaptive fuzzy rule-based classification system integrating both expert knowledge and data," in *Proc. 2012 ICTAI*, Athens, Greece, 2012, pp. 814–821.

[14] C. Tsang, S. Kwong, and H. Wang, "A systematic fuzzy rule based approach for fault classification in transmission lines," *Pattern Recogn.*, vol. 40, no. 2, pp. 2373–2391, 2007.

[15] A. Quteishat, C. P. Lim, and S. T. Kay, "A modified fuzzy min-max neural network with a genetic-algorithm-based rule extractor for pattern classification," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 40, no. 3, pp. 641–650, 2010.

[16] J. Sanz, A. Fernández, H. Bustince, and F. Herrera, "A genetic tuning to improve the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets: degree of ignorance and lateral position," *Int. J. Approx. Reason.*, vol. 52, no. 6, pp. 751–766, 2011.

[17] J. A. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, and H. Bustince, "Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system," *Appl. Soft Comput.*, vol. 20, no. 1, pp. 103–111, 2014.

[18] R. Lowen, *Fuzzy Set Theory: Basic Concepts, Techniques and Bibliography*. New York, NJ: Springer, 2012.

[19] A. Dempster, "Upper and lower probabilities induced by multivalued mapping," *Ann. Math. Statist.*, vol. 38, no. 2, pp. 325–339, 1967.

[20] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press, 1976.

[21] J. B. Yang, J. Liu, J. Wang, H. S. Sii, and H. W. Wang, "Belief rule-based inference methodology using the evidential reasoning approach–RIMER," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 36, no. 2, pp. 266–285, 2006.

[22] G. L. Kong, D. L. Xu, and B. Richard, "A belief rule-based decision support system for clinical risk assessment of cardiac chest pain," *Eur. J. Oper. Res.*, vol. 219, no. 3, pp. 564–573, 2012.

[23] B. Li, H. W. Wang, J. B. Yang, M. Guo, and C. Qi, "A belief-rule-based inventory control method under nonstationary and uncertain demand," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14 997–15 008, 2011.

[24] Z. J. Zhou, C. H. Hu, J. B. Yang, D. L. Xu, and D. H. Zhou, "Online updating belief rule based system for pipeline leak detection under expert intervention," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7700–7709, 2009.

[25] L. Jiao, Q. Pan, T. Denœux, Y. Liang, and X. Feng, "Belief rule-based classification system: Extension of FRBCS in belief functions framework," *Inform. Sciences*, vol. 309, no. 1, pp. 26–49, 2015.

[26] P. Smets, "Decision making in the TBM: the necessity of the pignistic transformation," *Int. J. Approx. Reason.*, vol. 38, no. 2, pp. 133–147, 2005.

[27] H. Helbig, *Knowledge Representation and the Semantics of Natural Language*. New York, NJ: Springer, 2006.

[28] O. Dieste and N. Juristo, "Systematic review and aggregation of empirical studies on elicitation techniques," *IEEE Trans. Software Eng.*, vol. 37, no. 2, pp. 283–304, 2011.

[29] O. Cordón, M. José del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systmes," *Int. J. Approx. Reason.*, vol. 20, no. 1, pp. 21–45, 1999.

[30] B. Ristic and P. Smets, "Target classification approach based on the belief function theory," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 2, pp. 574–583, 2005.

[31] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets Syst.*, vol. 52, no. 1, pp. 21–32, 1992.

[32] L. Jiao, T. Denœux, and Q. Pan, "Evidential editing k-nearest neighbor classifier," in *Lecture Notes in Computer Science*, S. Destercke and T. Denœux, Eds. Cham, Switzerland: Springer, 2015, vol. 9161, pp. 461–471.

[33] J. A. Sáez, M. Galar, J. Luengo, and F. Herrera, "Tackling the problem of classification with noisy data using multiple classifier systems: Analysis of the performance and robustness," *Inform. Sciences*, vol. 247, no. 1, pp. 1–20, 2013.

[34] C. Liang, Y. Zhang, and Q. Song, "Decision tree for dynamic and uncertain data streams," in *Proc. 2010 ACML*, Tokyo, Japan, 2010, pp. 209–224.

[35] S. Chen and H. He, "Towards incremental learning of nonstationary imbalanced data stream: a multiple selectively recursive approach," *Evol. Syst.*, vol. 2, no. 1, pp. 35–50, 2011.

[36] T. R. Hoens, R. Polikar, and N. V. Chawla, "Learning from streaming data with concept drift and imbalance: an overview," *Prog. Artif. Intell.*, vol. 1, no. 1, pp. 89–101, 2012.
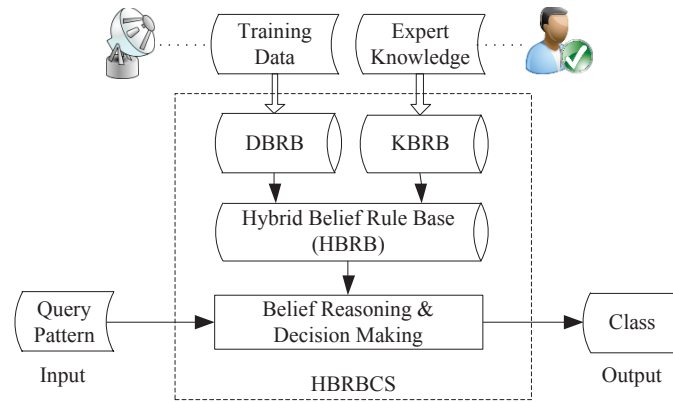
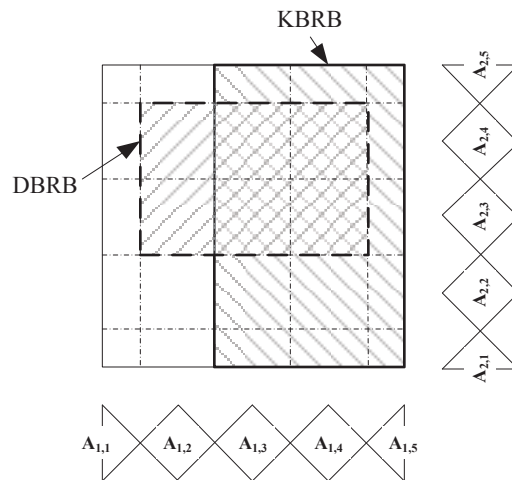Fig. 1. Hybrid belief rule-based classification system

Fig. 2. An example of the fuzzy regions covered by the DBRB and KBRB for a two-dimensional feature space
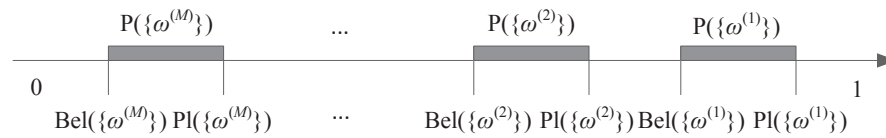

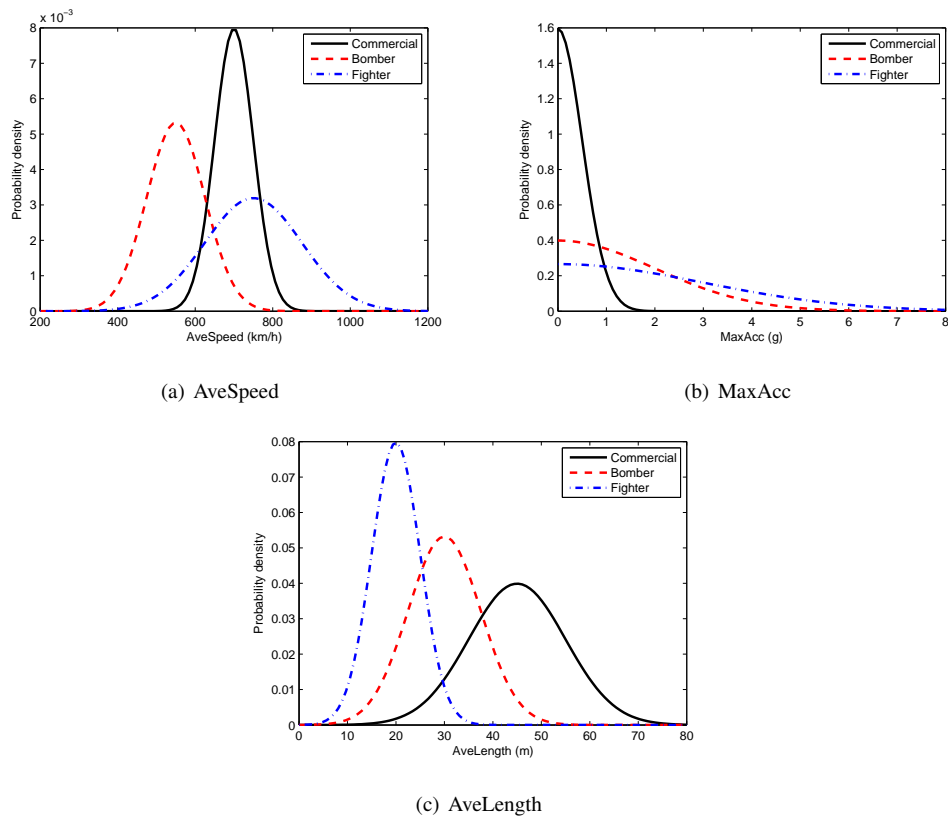Fig. 3. The credibility, plausibility, and probability for all of the M classes in the real interval $[0, 1]$

(a) AveSpeed

(b) MaxAcc

(c) AveLength

Fig. 4. Distributions of the three features conditioned on the class
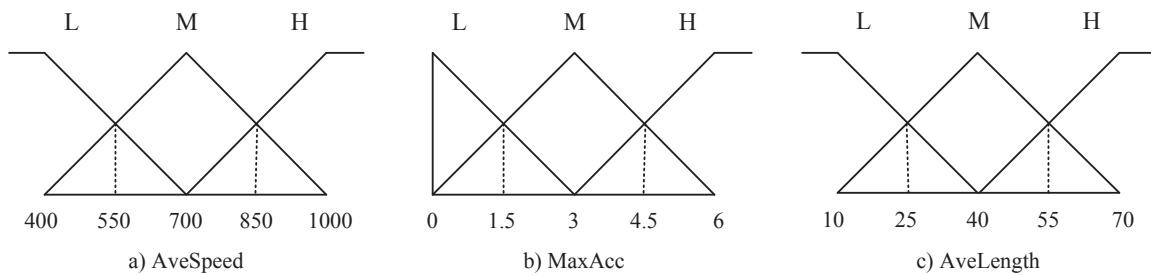

a) AveSpeed

b) MaxAcc

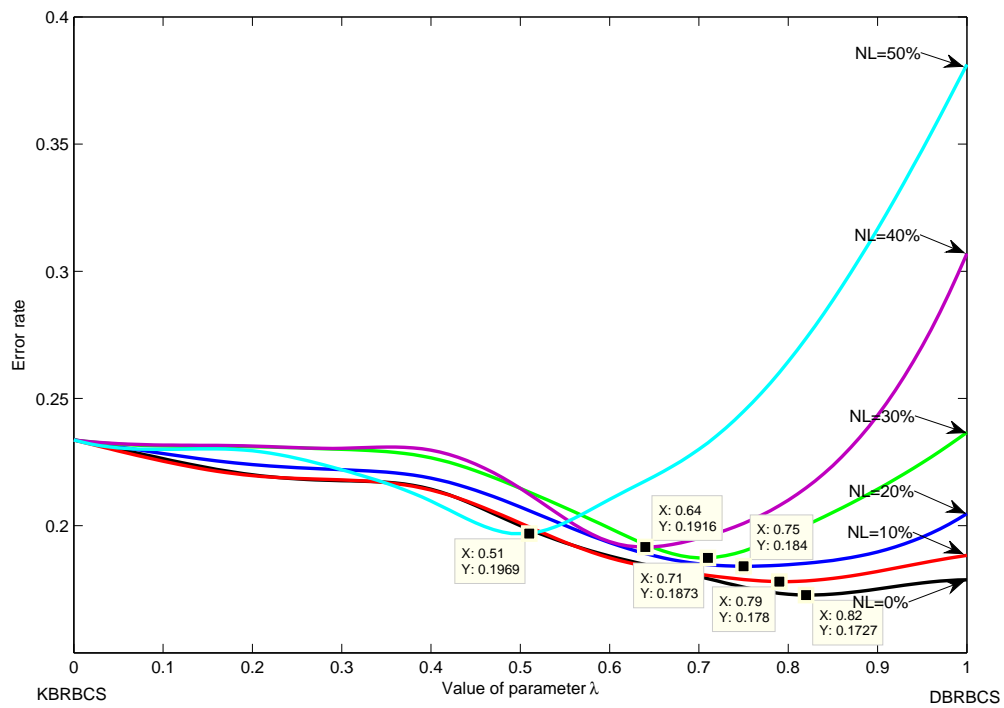c) AveLength

Fig. 5. Fuzzification of the feature space

Fig. 6.   Classification error rate of the HBRBCS with the adjustment factor ranging from 0 to 1