Expert Systems with Applications 42 (2015) 3410–3419

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# A weighted inference engine based on interval-valued fuzzy relational theory

Chee Kau Lim, Chee Seng Chan *

Centre of Image and Signal Processing, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Available online 26 December 2014

Keywords:
Fuzzy relations
BK subproduct
Inference engine
Interval-valued fuzzy sets
Fuzzy sets and systems
Membership functions

ABSTRACT

The study of fuzzy relations forms an important fundamental of fuzzy reasoning. Among all, the research on compositional fuzzy relations by Bandler and Kohout, or the Bandler–Kohout (BK) subproduct gained remarkable success in developing inference engines for numerous applications. Despite of its successfulness, we notice that there are limitations associated in the current implementations of the BK subproduct. In this paper, the BK subproduct, which originally based on the ordinary fuzzy set theory, is extended to the interval-valued fuzzy sets. This is because studies had claimed that ordinary fuzzy set theory has its limitation in addressing uncertainties using the crisp membership functions. Secondly, with the understanding that some features might have higher influence compare to the others, a weight parameter is introduced in the BK subproduct-based inference engines. Finally, a fuzzification method that able to fuzzify the input data and also train the inference engines is also developed. So, the BK subproduct-based inference systems can be built without human intervention, which are cumbersome and time consuming. Experiments on three public datasets and a comparison with state-of-art solutions have shown the efficiency and robustness of the proposed method.



## 1. Introduction

Reasoning with fuzzy sets theory has been widely studied in the literature. Some common directions of studies include finding soft cluster centers of data (Roh, Pedrycz, & Ahn, 2014; Son, 2015), integrate the fuzzy sets theory with other technologies such as artificial neural networks (Egrioglu, Aladag, & Yolcu, 2013; Suparta & Alhasa, 2014) or particle swarm optimization (Melin et al., 2013), modeling parameters as fuzzy numbers (Petrović et al., 2014; Samantra, Datta, & Mahapatra, 2014) and etc (Abdullah & Najib, 2014; Nguyen, Dawal, Nukman, & Aoyama, 2014). Among all, many researches focus on constructing fuzzy inference systems for various purposes, such as classification (Ait Laasri, Akhouayri, Agliz, Zonta, & Atmani, 2015; D'Andrea & Lazzerini, 2013; Samuel, Omisore, & Ojokoh, 2013; Yuste, Triviño, & Casilari, 2013), control (Bugarski, Bačkalić, & Kuzmanov, 2013; Liu, Han, & Lu, 2013; Liu, Yang, & Yang, 2013) and etc (Camastra et al., 2015; Gupta, Saini, & Saxena, 2014). In all these fuzzy inference systems, fuzzy rules are the core of the inference processes.

However, the popularity of these rule based fuzzy inference systems dispute the necessity of developing other inference schemes. Among all, a theory on compositions of fuzzy relations, namely the Bandler-Kohout (BK) subproduct (Kohout & Bandler, 1980a, 1980b) shows its excellency in some studies (Bodenhofer, Dankova, Stepnicka, & Novak, 2007; Stepnicka & Jayaram, 2010). Compare to those popular rule based fuzzy inference systems, a special characteristic of BK subproduct based inference systems is it does not require rules. Certainly, it is an advantage in the cases which rules are hardly define (Kolodner, 1992). While some theoretical researches on BK subproduct are in advancement (Mandal & Jayaram, 2013; Stepnicka & De Baets, 2013; Štěpnička & Holčapek, 2014), empirical works have also been carried out to prove its advantages in many expert systems, such as medical diagnostic systems (Kohout, Stabile, Kalantar, & San-Andres, 1995; Yew & Kohout, 1997), information retrieval (Kohout & Bandler, 1985), land evaluation (Groenemans, Ranst, & Kerre, 1997), cognitive sciences (Kohout, 2009), scene understanding (Vats, Lim, & Chan, 2012) and etc.

Despite of its successfulness, we notice that there are still some limitations associated in the current implementations of the BK subproduct. First of all, the implementations of the BK subproduct in the literature are still based on the classical Type-1 Fuzzy Set (T1FS) theory, which address uncertainties with point-values. Studies such as Mendel (2000, 2003) claimed that T1FS has its limitation in addressing uncertainties with its crisp membership functions. Secondly, the BK subproduct performs inferences utilize a set

* Corresponding author. Tel.: +60 3 7967 6433.
E-mail addresses: limck@siswa.um.edu.my (C.K. Lim), cs.chan@um.edu.my (C.S. Chan).

http://dx.doi.org/10.1016/j.eswa.2014.12.025
0957-4174/© 2014 Elsevier Ltd. All rights reserved.

of common features that relate the inputs and outputs. In most cases, the BK subproduct treats all the features equally, i.e. the importance of all the features are similar. However, in practical, not all these features have the identical influences towards the inference outcomes. We argue that some features may have higher reliability or distinguishability than the others, and vice versa. In the literature of fuzzy logic, implementation of the weight parameter is not rare (Ishibuchi & Yamamoto, 2005; Seki & Mizumoto, 2011; Xing & Ha, 2014). For example, Groenemans et al. (1997) have tried to incorporate the weight parameter in the BK subproduct. However, this implementation would require to fulfill a condition: $\sum_{n=1}^{N} w_n = 1$ where $n = \{1, 2, \ldots, N\}, N$ is the number of features and $w_n$ is the weight of feature $n$. This condition is too restrictive for a good implementation of weight because: *(i)* adding or decreasing features into consideration list will cause recalculation of all the weights. For instance, adding a new feature with weight $w_{N+1} \neq 0$ to the existing feature list will cause the total weight become $\sum_{n=1}^{N+1} w_n > 1$ and the condition of total weight equal to 1 is not longer hold. Thus, a normalization is required so that the $\sum_{n=1}^{N+1} w_n = 1$ is satisfied. *(ii)* importance or influence of a feature is not intuitive – i.e. comparing a system with such condition to an implementation of weight where $w_n \in [0, 1]$ for all $n$, the weights of the latter are much intuitive as the weights close to 0 means less influence, while close to 1 means high influence. Groenemans et al. (1997) found out that the weights can be small numbers that close to 0 even if they have high influence in the case of number of features $N$ is large. Furthermore, this problem become much more complicated if new features that are going to be added into consideration, as one may not know what are the appropriate values that representing the high (or low) influence.

Last but not least, in all the previous attempts of the BK subproduct, predefine rules (Bui & Kim, 2006) or experts knowledge (Kohout et al., 1995) are priori required, so that the knowledge base can be formed. In some cases, worst still the fuzzification of the input data are done manually, which is cumbersome and time consuming. An approach to train the BK subproduct automatically so that it can learn from the training examples could not be found in the literature, and the lack of this learning mechanism greatly limits the application of the BK subproduct in many fields.

Hence, the aims of this paper is to form a more reliable BK subproduct reasoning framework. For this purpose, our contributions are threefolds: (1) we extend the current BK subproduct to the Interval-Valued Fuzzy Sets (IVFS) by defining a subsethood measure of IVFS. (2) a weight parameter is incorporated to the BK subproduct-based inference engines so that more attention is given to those important features, and (3) we introduce a learning mechanism for the BK subproduct. Employing the training samples, the proposed learning mechanism manages to form the knowledge base of a BK subproduct-based inference engine without human intervention. Additionally, the learning mechanism also produce membership functions that serve to fuzzify the input data. To prove the advantages and improvements of the proposed BK subproduct, three publicly available medical data sets are employed. Experimental results and a comparison with state-of-the-art solutions have shown the efficiency of the proposed method.

The rest of the paper are arranged as follow: In Section 2, we provide a short revision on the BK subproduct. Section 3 discusses the extension of BK subproduct from T1FS to IVFS, along with a subsethood measure of IVFS. In Section 4, we introduce the weight parameter to this newly developed IVFS reasoning scheme. Section 5 proposes a learning mechanism so that the BK subproduct based inference systems can be built from numerical data. The classification experiment and the discussion is presented in Section 6. Lastly, we conclude the paper in Section 7.

## 2. BK subproduct revisit

We start this section with a brief review on the definition of the BK subproduct on crisp relations. Assume that $A, B$ and $C$ are three crisp sets and $a, b$ and $c$ are general representation to the elements in these sets respectively. $R$ is defined as a relation from $A$ to $B$ such that $R \subseteq A \times B$; whereas $S$ is a relation from $B$ to $C$ such that $S \subseteq B \times C$. The converse relation of $S$ is denoted as $S^T$. The abbreviation $aRb$ shows that $a$ is in relation $R$ with $b$. Kohout and Bandler (1980a) defined the BK subproduct as follow:

**Definition 1.** The BK subproduct is a composition of relations for $a$ and $c$ such that:

$$R \triangleleft S = \{(a, c) | (a, c) \in A \times C \quad \text{and} \quad aR \subseteq Sc\} \tag{1}$$

where $aR = \{b | aRb\}$ is the image of $a$ after the projection of relation $R$ in the set $B$, while $Sc = \{b | bSc\}$ is the image of $c$ after the projection of relation $S^T$ in the set $B$. The BK subproduct is useful in retrieving relationships between elements of two indirectly associated sets, *objects* $A$ and *targets* $C$, if both sets can be associated with a set of common *features*, $B$.

It is trivial that Definition 1 is established on the subsethood of $aR$ in $Sc$. In order to extend it to the fuzzy relations, the fuzzy subsethood measure for T1FS is defined as:

**Definition 2.** For two T1FS, $P$ and $Q$ in the same universe $X$, the possibility of $P$ is a subset of $Q$ is given as follow:

$$\pi(P \subseteq Q) = \bigwedge_{x \in X} (P(x) \rightarrow Q(x)) \tag{2}$$

where $P(x)$ and $Q(x)$ are the membership degrees of $x$ in set $P$ and $Q$ respectively, $\rightarrow$ is the fuzzy implication operators (see Table 1) generally defined as "NOT $A$ OR $B$" and $\wedge$ is the infimum operator which can be considered as min function in harsh criterion or arithmetic mean in mean criterion (Kohout & Bandler, 1980a).

**Definition 3.** Employing the Eqs. (1) and (2), Kohout and Bandler (1980a) defined the fuzzy BK subproduct as follow:

$$R \triangleleft S(a, c) = \bigwedge_{b \in B} (R(a, b) \rightarrow S(b, c)) \tag{3}$$

where $R(a, b)$ is the fuzzy membership degree to which $aRb$ is true and $S(b, c)$ is the fuzzy membership degree to which $bSc$ is true.

**Table 1**
Example of the fuzzy implication operators and their respective definitions.

| Name | Symbol | Definition |
|---|---|---|
| S# – Standard Sharp (Mizumoto & Zimmermann, 1982) | $r \rightarrow_{S\#} s$ | $\begin{cases} 1 & \text{iff } r \neq 1 \text{ or } s = 1 \\ 0 & \text{otherwise} \end{cases}$ |
| S – Standard Strict (Mizumoto & Zimmermann, 1982) | $r \rightarrow_{S} s$ | $\begin{cases} 1 & \text{iff } r \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$ |
| G43 – Gaines 43 (Mizumoto & Zimmermann, 1982) | $r \rightarrow_{G43} s$ | $\min(1, \frac{r}{s})$ |
| KD – Kleene-Dienes (Kohout & Bandler, 1980a) | $r \rightarrow_{KD} s$ | $\max(s, 1 - r)$ |
| R – Reichenbach (Kohout & Bandler, 1980a) | $r \rightarrow_{R} s$ | $1 - r + rs$ |
| L – Łukasiewicz (Zadeh, 1975) | $r \rightarrow_{Ł} s$ | $\min(1, 1 - r + s)$ |
| Y – Yager (Yager, 1980) | $r \rightarrow_{Y} s$ | $s^r$ |
| EZ – Early Zadeh (Zadeh, 1975) | $r \rightarrow_{EZ} s$ | $(r \wedge s) \vee (1 - r)$ |

## 3. Interval-valued fuzzy sets-based BK subproduct

The conventional BK subproduct are based on the fuzzy subsethood measure that relies on fuzzy implication operators. Therefore it is not suitable for proposed IVFS-based BK subproduct as all the fuzzy implication operators are only defined for point-value membership degrees, and the membership degrees for IVFS are intervals. To tackle this problem, in this paper, we proposed an IVFS subsethood measurement method based on the implication operators, namely the Complete Derivation Method.

Let $\tilde{P}$ and $\tilde{Q}$ be the two IVFS in the universe $X$ as shown in Fig. 1. If $I$ is the number of elements in the universe, for an element $x_i, i = \{1, \ldots, I\}$, the interval-valued membership degrees for $x_i$ in $\tilde{P}$ and $\tilde{Q}$ are $[\underline{P}(x_i), \overline{P}(x_i)]$ and $[\underline{Q}(x_i), \overline{Q}(x_i)]$ respectively. Assume that both axes of element and membership degrees are discrete, or can be discretized, Representation Theorem (Mendel & John, 2002) suggests that, if the total number of T1FS for $\tilde{P}$ is $\eta_{\tilde{P}}$, then the number of T1FS that passing through a discrete point $P(x_i)_j$ is given by $\frac{\eta_{\tilde{P}}}{J_i}$, where $J_i$ is the total number of discrete membership degrees in $[\underline{P}(x_i), \overline{P}(x_i)]$ and $j = \{1, \ldots, J_i\}$. For $\tilde{Q}$, if $K_i$ is the number of discrete membership degree in $[\underline{Q}(x_i), \overline{Q}(x_i)]$ and $k = \{1, 2, \ldots, K_i\}$, the number of T1FS that passing through a discrete point $Q(x_i)_k$ is given by $\frac{\eta_{\tilde{Q}}}{K_i}$.

To formulate the fuzzy subsethood measure of IVFS, we start with evaluating an arbitrary pair of point membership degrees $P(x_i)_j$ and $Q(x_i)_k$ in $\tilde{P}$ and $\tilde{Q}$ respectively, on a same element $x_i$. If these are the only points on $x_i$, following Eq. (2), the subsethood measure on this element can be written as $\pi(\tilde{P} \subseteq \tilde{Q})(x_i) = P(x_i)_j \to Q(x_i)_k$. However, since there are $\frac{\eta_{\tilde{P}}}{J_i}$ of T1FS on $P(x_i)_j$ and $\frac{\eta_{\tilde{Q}}}{K_i}$ on $Q(x_i)_k$, this implication involves a number of $\frac{\eta_{\tilde{P}}}{J_i} \times \frac{\eta_{\tilde{Q}}}{K_i}$ pairs of T1FS, thus it should be represented as:

$$\frac{\eta_{\tilde{P}}\eta_{\tilde{Q}}}{J_i K_i}\left(P(x_i)_j \to Q(x_i)_k\right) \tag{4}$$

If $P(x_i)_j$ is the only discrete point for $\tilde{P}(x_i)$, then the subsethood measure can be evaluate by summing up all the implications of this point membership degree to all the $Q(x_i)_k$ as illustrated in Fig. 2):

$$\frac{\eta_{\tilde{P}}\eta_{\tilde{Q}}}{J_i K_i}\sum_{k=1}^{K_i}\left(P(x_i)_j \to Q(x_i)_k\right) \tag{5}$$

In general cases, $[\underline{P}(x_i), \overline{P}(x_i)]$ are intervals with more than one discrete points. Therefore, if we generalized Eq. (5) to all the $P(x_i)_j$ in the element $x_i$ and normalized it with the total number of T1FS $\eta_{\tilde{P}}\eta_{\tilde{Q}}$, we have the fuzzy subsethood measure for this element as:

$$\pi(\tilde{P} \subseteq \tilde{Q})(x_i) = \frac{1}{J_i K_i}\sum_{j=1}^{J_i}\sum_{k=1}^{K_i}\left(P(x_i)_j \to Q(x_i)_k\right) \tag{6}$$

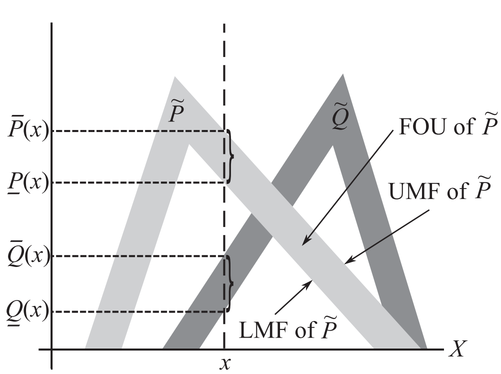


**Fig. 1.** Two IVFS $\tilde{P}$ and $\tilde{Q}$ in the same universe $X$.

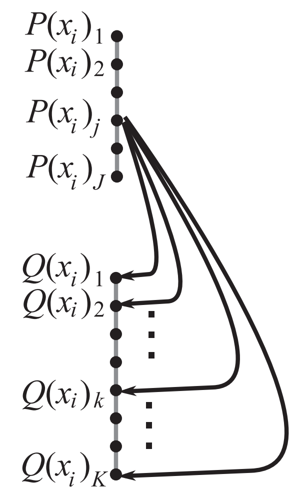


**Fig. 2.** Implication of a single membership degree in $\tilde{P}(x_i)$ to all the discretized membership degrees of $\tilde{Q}(x_i)$.

Substitute Eq. (6) to Eq. (3), the IVFS subsethood measure extended from the original BK subproduct is:

$$\pi(\tilde{P} \subseteq \tilde{Q}) = \bigwedge_{i \in I}\frac{1}{J_i K_i}\sum_{j=1}^{J_i}\sum_{k=1}^{K_i}\left(P(x_i)_j \to Q(x_i)_k\right) \tag{7}$$

With a fuzzy implication operator, Eq. (7) will result a subsethood measurement of $\tilde{P} \subseteq \tilde{Q}$ in the interval $[0, 1]$. However, one must note that the subsethood measurements for crisp sets are Boolean (yes/no) and for the T1FS are point-values. Therefore, it is reasonable to deduce that the subsethood measurements for the IVFS should be in intervals instead of point-values Nguyen and Kreinovich (2008), Yang and Lin (2009), Zheng, Xiao, Zhang, and Shi (2010) and Rickard, Aisbett, and Greb (2009). Studies from Kohout and Bandler (1980b, 1992) and Lim and Chan (2011, 2012) suggested that the Kleene-Dienes and Łukasiewicz implication operators are amongst the two most suitable candidates for lower and upper bounds of fuzzy subsethood measurements. With the knowledge that:

$$p \to_{\mathrm{KD}} q \leqslant p \to_{Ł} q \quad \forall p, q \in [0, 1] \tag{8}$$

Thus, the subsethood measure using the Complete Derivation Method is given by:

$$\pi(\tilde{P} \subseteq \tilde{Q}) = \left[\bigwedge_{i \in I}\frac{1}{J_i K_i}\sum_{j=1}^{J_i}\sum_{k=1}^{K_i}\left(P(x_i)_j \to_{\mathrm{KD}} Q(x_i)_k\right), \bigwedge_{i \in I}\frac{1}{J_i K_i}\sum_{j=1}^{J_i}\sum_{k=1}^{K_i}\left(P(x_i)_j \to_{Ł} Q(x_i)_k\right)\right] \tag{9}$$

Assume that the $\tilde{P}(x)$ is the image of $a\tilde{R}$ and $\tilde{s}c$ gives $\tilde{Q}(x)$, the proposed IVFS-based BK subproduct can be defined as follow:

$$\tilde{R} \triangleleft \tilde{s}(a, c) = \left[\frac{1}{I}\sum_{i=1}^{I}\frac{1}{J_i K_i}\sum_{j=1}^{J_i}\sum_{k=1}^{K_i}\max\left(S(b_{ik}, c), 1 - R(a, b_{ij})\right), \frac{1}{I}\sum_{i=1}^{I}\frac{1}{J_i K_i}\sum_{j=1}^{J_i}\sum_{k=1}^{K_i}\min\left(1, 1 - R(a, b_{ij}) + S(b_{ik}, c)\right)\right] \tag{10}$$

## 4. Weighted IVFS-based BK subproduct

### 4.1. Weight and BK subproduct

Generally, we cluster criteria into a few criteria sets during the reasoning process. Among these criteria sets, some of them might

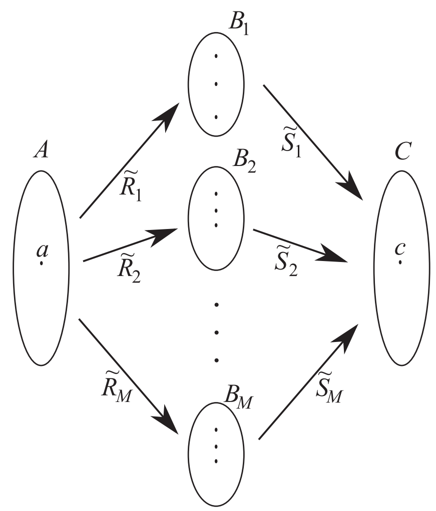


**Fig. 3.** Breaking up set $B$ into multiple feature sets to form weighted BK subproduct.

have higher influence over the others in a decision making process. Hence, weights are very useful parameter to represent the influence of the criteria sets. However, one should note that the weight should not be confused with the strength of criteria in the criteria set. Similar case is applied to the BK subproduct-based inference engines where instead of criteria, features are being employed. Membership degrees of the object-feature relations ($\tilde{R}$) and feature-target relations ($\tilde{s}$) are the "strength of criteria" that determine the inference outcome; while features from the feature sets, the influence of each of the feature sets are represented as weight. That is, weight is applied to each feature set and is modeled in the IVFS.

As mentioned earlier, the subsethood measure is the fundamental of fuzzy BK subproduct. Thus, one might argue that it is inappropriate to implement weight in the BK subproduct based inference engines because there is no well defined weighted subsethood measure in the literature. In fact, the weights are applied to the feature sets rather than the subsethood measurements. We explain this argument with a multiple feature sets model accordingly.

Assume that the features in set $B$ can be clustered into multiple feature sets $B_m, m = \{1, 2, \ldots, M\}$ and each of the feature set has a number of features. The relation between $A$ and $B_m$ is $\tilde{R}_m$, whereas $\tilde{s}_m$ is the relation between $B_m$ and $C$. In this case, the images of $a\tilde{R}_m$ and $\tilde{s}_m c$ are $\tilde{P}_m$ and $\tilde{Q}_m$, respectively. Studying the subsethood measure of $\tilde{P}_m \subseteq \tilde{Q}_m$, one can get the BK subproduct $\tilde{R}_m \lhd \tilde{s}_m(a,c)$ (Fig. 3) and each of the feature set carries different weights. Assume that the weight of $\tilde{R}_m \lhd \tilde{s}_m(a,c)$ is $\widetilde{W}_m$, the normalized aggregation of all the composition of relations is given as:

$$\tilde{R} \lhd \tilde{s}(a,c) = \frac{\sum_{m=1}^{M} \widetilde{W}_m \left(\tilde{R}_m \lhd \tilde{s}_m(a,c)\right)}{\sum_{m=1}^{M} \widetilde{W}_m} \tag{11}$$

Eq. (11) gives the weighted measure of the IVFS-based BK subproduct. Here, since all $\tilde{R}_m \lhd \tilde{s}_m(a,c)$ are intervals that only exist as numerators, whereas $\widetilde{W}_m$ are IVFS, the computations results based on Eq. (11) are always IVFS. We show the details of solving Eq. (11) in the following subsection.

### 4.2. Solving the weighted average

Solving Eq. (11) is easy if all the parameters are crisp numbers. However, in this paper, these parameters are fuzzy, and so the solution become slightly complicated, especially with a term $1/\sum_{m=1}^{M} \widetilde{W}_m$. One of the closest solution is the Fuzzy Weighted Average (FWA) (Dong & Wong, 1987). The FWA solved the problems in the form of:

$$f = \sum_{m=1}^{M} (\omega_m \chi_m) \Big/ \sum_{m=1}^{M} \omega_m \tag{12}$$

where all $\chi_m$ and $\omega_m$ are T1FS. Wu and Mendel (2007) extended the FWA to form Linguistic Weighted Average (LWA), where all $\chi_m$ and $\omega_m$ are IVFS. Both the FWA and LWA use α-cut decomposition theorem (Klir & Yuan, 1995). Using the α-cut decomposition theorem, instead of performing computations directly on the sets ($\chi_m$ and $\omega_m$) as whole, a number of $(\delta - 1)$ α-cuts are taken to break the sets into $\delta$ intervals. For each interval $\mathcal{I}_\imath, 1 \leqslant \imath \leqslant \delta$, perform computation on the obtained intervals after the α-cut, i.e. $\chi_m^\imath$ and $\omega_m^\imath$ to yield an interval $\mathfrak{I}_\imath$. The composition of all the $\mathfrak{I}_\imath$ with corresponding α-cuts form the corresponding set $\mathfrak{I}$.

In this paper, we engaged the solution of LWA proposed by Wu and Mendel (2007, 2008) with assumption that the intervals $\tilde{R}_m \lhd \tilde{s}_m(a,c)$ in Eq. (11) are special cases of the IVFS, that are these fuzzy sets have rectangle membership functions and the Upper Membership Functions (UMF) and Lower Membership Functions (LMF) of $\tilde{R}_m \lhd \tilde{s}_m(a,c)$ are equal. From here onwards, we denote $\tilde{R}_m \lhd \tilde{s}_m(a,c)$ as $Z_m$ and the lower and upper bounds of $Z_m$ are denoted as $\underline{Z}_m$ and $\overline{Z}_m$ respectively. Thus, following this notation scheme, $\tilde{R} \lhd \tilde{s}(a,c)$ is denoted as $\widetilde{Z}$.

Since the Footprint of Uncertainty (FOU) of $\widetilde{Z}$ is determined by UMF($\widetilde{Z}$) and LMF($\widetilde{Z}$), we can find $\widetilde{Z}$ by calculating these two boundaries only. Wu and Mendel (2008) proved that the height of the output sets from LWA are equal to the minimum height of all $Z_m$ and $W_m$. Herein, since all UMF($\widetilde{W}_m$) are normal and $Z_m$ are intervals, the height of an UMF($\widetilde{Z}$) is unity. On the other hand, the height of a LMF($\widetilde{Z}$) is totally depends on LMF($\widetilde{W}_m$). Assume that all the $\widetilde{W}_m$ have trapezoidal (or triangular) shape FOU, the shape of $\widetilde{Z}$ should be trapezoidal (or triangular) as well (Fig. 4).

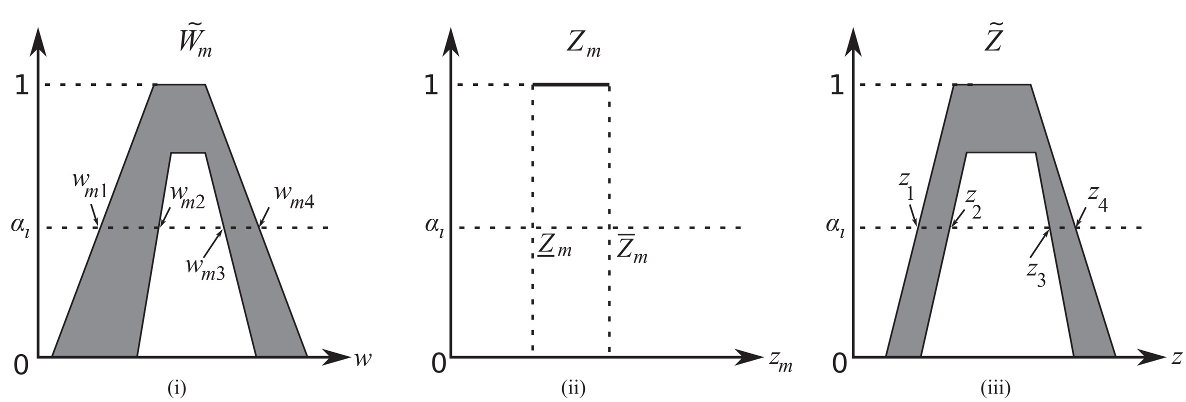


**Fig. 4.** Notations used in finding weighted results of IVFS-based BK subproduct.

As described earlier, the solution of Eq. (11) starts with taking $(\delta - 1)$ $\alpha$-cuts to yield $\delta$ intervals for each set. Thus, the rest of the work is simplified to find the intervals that represent the FOU of $\widetilde{Z}$ corresponding to each $\alpha$-cut. For this purpose, notations that described in Fig. 4 are used:

(i) $\widetilde{W}_m$ : for an $\alpha$-cut $\alpha_\imath, w_{\imath m1}$ and $w_{\imath m4}$ should be the leftmost and rightmost values of UMF($\widetilde{W}_m$) respectively at $\alpha_\imath$. However, the variable $\imath$ is intentionally left out as an subscript of all variables here because it is independent from the calculation of each iteration, and to make the equations look more concise. Therefore, these variables become $w_{m1}$ and $w_{m4}$. Similarly, $w_{m2}$ and $w_{m3}$ are the leftmost and rightmost values of LMF($\widetilde{W}_m$) respectively.
(ii) $Z_m$ : $\underline{Z}_m$ is the lower bound of interval $Z_m$, whereas $\overline{Z}_m$ is the upper bound of this interval.
(iii) $\widetilde{Z}$ : for an $\alpha$-cut $\alpha_\imath, z_1$ and $z_4$ are the leftmost and rightmost values of UMF($\widetilde{Z}$) respectively. Similarly, $z_2$ and $z_3$ are the leftmost and rightmost values of LMF($\widetilde{Z}$) respectively.

Refer to the results of LWA (Wu & Mendel, 2007, 2008), for each $\alpha$-cut, the corresponding boundaries of UMF($\widetilde{Z}$) and LMF($\widetilde{Z}$) can be obtained by sorting $\overline{Z}_m$ and $\underline{Z}_m$ in ascending order, then substituting the corresponding values into the following equations:

$$z_1 = \frac{\sum_{m=1}^{\beta_1} w_{m4}\underline{Z}_m + \sum_{m=\beta_1+1}^{M} w_{m1}\underline{Z}_m}{\sum_{m=1}^{\beta_1} w_{m4} + \sum_{m=\beta_1+1}^{M} w_{m1}} \tag{13}$$

$$z_2 = \frac{\sum_{m=1}^{\beta_2} w_{m3}\underline{Z}_m + \sum_{m=\beta_2+1}^{M} w_{m2}\underline{Z}_m}{\sum_{m=1}^{\beta_2} w_{m3} + \sum_{m=\beta_2+1}^{M} w_{m2}} \tag{14}$$

$$z_3 = \frac{\sum_{m=1}^{\beta_3} w_{m2}\overline{Z}_m + \sum_{m=\beta_3+1}^{M} w_{m3}\overline{Z}_m}{\sum_{m=1}^{\beta_3} w_{m2} + \sum_{m=\beta_3+1}^{M} w_{m3}} \tag{15}$$

$$z_4 = \frac{\sum_{m=1}^{\beta_4} w_{m1}\overline{Z}_m + \sum_{m=\beta_4+1}^{M} w_{m4}\overline{Z}_m}{\sum_{m=1}^{\beta_4} w_{m1} + \sum_{m=\beta_4+1}^{M} w_{m4}} \tag{16}$$

In these equations, $\beta_1, \beta_2, \beta_3$ and $\beta_4$ are the switching points in the range $[1, M]$ calculated using the Karnik–Mendel algorithm (Liu & Mendel, 2008; Mendel, 2009) such that:

$$\underline{Z}_{\beta_1} \leqslant z_1 \leqslant \underline{Z}_{\beta_1+1} \tag{17}$$
$$\underline{Z}_{\beta_2} \leqslant z_2 \leqslant \underline{Z}_{\beta_2+1} \tag{18}$$
$$\overline{Z}_{\beta_3} \leqslant z_3 \leqslant \overline{Z}_{\beta_3+1} \tag{19}$$
$$\overline{Z}_{\beta_4} \leqslant z_4 \leqslant \overline{Z}_{\beta_4+1} \tag{20}$$

## 5. Learning mechanism for IVFS-based BK-subproduct

A typical fuzzy rule based system (Fazel Zarandi, Rezaee, Turksen, & Neshat, 2009; Lam & Seneviratne, 2008; Lee, 1990a) has rules in the form:

if $\mathsf{X}_1$ is $U_1$ and $\mathsf{X}_2$ is $U_2$ then $\mathsf{Y}$ is $V$

Here, $\mathsf{X}_1$ and $\mathsf{X}_2$ are antecedents limited by linguistic terms $U_1$ and $U_2$ respectively, whereas $\mathsf{Y}$ is the consequent that related to another linguistic terms $V$. $U_1$ and $U_2$ can be in different universe, as well as the $V$.

In contrast, for the BK subproduct based inference engines, there is no rules exist and inferences are only based on the relations of $\tilde{R}(a, b)$ and $\tilde{s}(b, c)$, which both come from the same domain for a feature $b$. Therefore, finding the values of $\tilde{R}(a, b)$ and $\tilde{s}(b, c)$ are interrelated and are in accordance with the domain of $b$. Moreover, the construction of membership is independence for each feature $b$. Since $\tilde{R}(a, b)$ are mappings from $a$ to $b$ and $\tilde{s}(b, c)$ are reverse mapping from $c$ to $b$, the domain of $b$ should be studied. Assume that the number of features in set $B$ is $J \in \mathbb{N}$ and $j = \{1, \dots, J\}$. Each $a_i \in A$, where $i \in \{1, \dots, I\}, I \in \mathbb{N}$, can be mapped to a feature $b_j$ with a value $L_{ji}$. With all the $L_{ji}$, the domain of the feature $b_j$ can be defined as an interval $L_j^{\mathrm{I}}$:

$$L_j^{\mathrm{I}} = \{L_{j1}, \dots, L_{jI}\} = [\underline{L}_j, \overline{L}_j] \tag{21}$$

Divide this domain into multiple sections, where the number of section is depends on the number of linguistic terms one want to define for the feature. These numbers can be different for different features. For each section, form a interval-valued membership function. There is no rules on the shape of the membership functions, but for the sake of simplicity, the membership functions with trapezoidal UMF and triangular LMF are used in the following. Assume that all the UMF are normal and the heights of LMF are $v$.

Let $H_j \in \mathbb{N}$ be the number of membership functions defined in the domain of feature $b_j$, and $h = \{1, \dots, H_j\}$. A membership function defined in this domain can be named as $\tilde{F}_{jh_j}$. Fig. 5 shows an example where domain of $b_j$ is divided into $H_j$ membership functions that represent linguistic terms "very low", "low", "high", "medium low", "very high" and etc.

For a membership function $\tilde{F}_{jh_j}$ that defined for a section, if the shape of this normal membership function is trapezoidal UMF and triangular LMF, one can define the parameters of this membership function as shown in Fig. 6. Parameters 1–4 define the UMF, parameters 5–7 and $v$ define the LMF. Assume that there are $K \in \mathbb{N}$ targets in $C$ and $k = \{1, \dots, K\}$. The composition of relation $\tilde{R} \triangleleft \tilde{s}(a_i, c_k)$ is meaningful only if $a_i$ implies $c_k$. Therefore, to find $\tilde{s}(b, c_k), A$ is partitioned into $K$ subsets according to $c_k$:

$$A = \{A_1, \dots, A_k, \dots, A_K\} \tag{22}$$
$$a_i^k \in A_k \iff a_i^k \to c_k \tag{23}$$

If $a_i^k$ maps to $b_j$ with a value $L_{ji}$, as proposed by Zadeh (1978) and Civanlar and Trussell (1986), the probability density function of $L_{ji}$

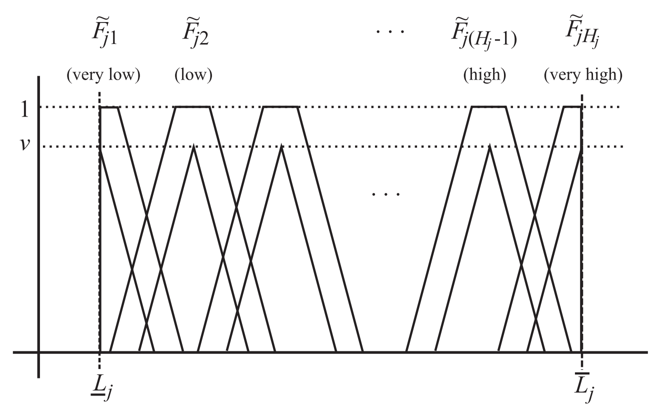


**Fig. 5.** Define a number of $H_j$ membership functions in the interval $L_j^{\mathrm{I}}$.

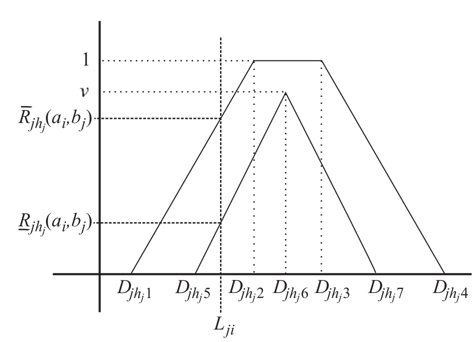


**Fig. 6.** The 8 points definition of a membership function of an IVFS and the mapping of membership degrees.

in $[\underline{L}_j, \overline{L}_j]$ can be studied to find the membership degrees of $\tilde{s}(b, c_k)$. Therefore, let $|A_k|$ be the number of elements in $A_k$, and $|A^{(5,7)}_{kjh_j}|$ be the number of elements of $a^k_i$ map to $b_j$ with a value $L_{ij}$ such that $D_{hj_h5} \leqslant L_{ij} \leqslant D_{hj_h7}$. With the information on the distribution of $a^k_i$ in the lower membership functions range, the probability density functions can be plotted. Following, $\underline{S}_{jh_j}(b_j, c_k)$ can be find:

$$\underline{S}_{jh_j}(b_j, c_k) = \frac{|A^{(5,7)}_{kjh_j}|}{|A_k|} \tag{24}$$

Similarly, if $|A^{(1,4)}_{kjh_j}|$ is the number of elements of $a^k_i$ map to $b_j$ with a value $L_{ij}$ in the upper membership functions range $[D_{hj_h1}, D_{hj_h4}]$, the upper bound of $\tilde{s}_{jh_j}(b_j, c_k)$ can be find:

$$\overline{S}_{jh_j}(b_j, c_k) = \frac{|A^{(1,4)}_{kjh_j}|}{|A_k|} \tag{25}$$

The Eqs. (24) and (25) find the relations between elements in set $B$ and set $C$, which are essential during the training process of a classifier. However, to prepare the inference engines for prediction, the testing data should be fuzzified in the fuzzification module to form $\tilde{R}(a, b)$. With the membership functions defined, finding the $\tilde{R}(a, b)$ is straight forward. Firstly, $\tilde{F}_{jh_j}$, the set of membership functions developed to find $\tilde{s}(b, c)$ must be adopted so that both $\tilde{s}(b, c)$ and $\tilde{R}(a, b)$ refer to the same set of membership functions. Subsequently, mapping of values described below gives the membership degrees of $\tilde{R}(a, b)$.

With relation $\tilde{R}(a_i, b_j)$, $a_i$ may maps to $b_j$ with a value $L_{ji}$ in the interval $L^{\mathrm{I}}_j$. If this $L_{ji}$ falls into the section where $\tilde{F}_{jh_j}$ defines (i.e. in $[D_{jh_j1}, D_{jh_j4}]$), we can retrieve a membership degree for this membership function, $\tilde{R}_{jh_j}(a_i, b_j)$, otherwise $\tilde{R}_{jh_j}(a_i, b_j) = 0$ for this membership function (Fig. 6). The upper and lower bound of this interval membership function, $[\underline{R}_{jh_j}(a_i, b_j), \overline{R}_{jh_j}(a_i, b_j)]$, is given by:

$$\underline{R}_{jh_j}(a_i, b_j) = \begin{cases} \frac{(v)(L_{ji}-D_{jh_j5})}{D_{jh_j6}-D_{jh_j5}} & \text{if } D_{jh_j5} < L_{ji} \leqslant D_{jh_j6} \\ \frac{(v)(D_{jh_j7}-L_{ji})}{D_{jh_j7}-D_{jh_j6}} & \text{if } D_{jh_j6} < L_{ji} < D_{jh_j7} \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

$$\overline{R}_{jh_j}(a_i, b_j) = \begin{cases} \frac{L_{ji}-D_{jh_j1}}{D_{jh_j2}-D_{jh_j1}} & \text{if } D_{jh_j1} < L_{ji} < D_{jh_j2} \\ 1 & \text{if } D_{jh_j2} \leqslant L_{ji} \leqslant D_{jh_j3} \\ \frac{D_{jh_j4}-L_{ji}}{D_{jh_j4}-D_{jh_j3}} & \text{if } D_{jh_j3} < L_{ji} < D_{jh_j4} \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

It is possible that in the testing data set, there are some cases that $a_i$ maps to $L'_{ji}$ where $L\prime_{ji} < \underline{L}_j$. In such cases, it is wiser to reconsider both the membership degrees of $\underline{R}_{jh_j}(a_i, b_j)$ and $\overline{R}_{jh_j}(a_i, b_j)$ if $\tilde{F}_{jh_j}$ is a left-shoulder membership function. For the case where $D_{jh_j1} = D_{jh_j2} = D_{jh_j5} = D_{jh_j6}$, one should set $\underline{R}_{jh_j}(a_i, b_j)$ and $\overline{R}_{jh_j}(a_i, b_j)$ to the heights of the corresponding membership functions, i.e. $v$ and 1 respectively. It is similar for the case when $L\prime_{ji} > \overline{L}_j$. For right shoulders membership functions $\tilde{F}_{jh_j}$ such that $D_{jh_j3} = D_{jh_j4} = D_{jh_j6} = D_{jh_j7}$, one should also set the membership degrees of both $\underline{R}_{jh_j}(a_i, b_j)$ and $\overline{R}_{jh_j}(a_i, b_j)$ to $v$ and 1 respectively, if they are the heights of the corresponding membership functions.

As a summary, this learning method forms $H_j$ membership functions for a feature $b_j$, thus, for a $c_k$, it finds $\sum_{j=1}^{J} H_j$ membership degrees for both $\underline{S}(b, c_k)$ and $\overline{S}(b, c_k)$. The total number of membership degrees of both $\underline{S}(b, c)$ and $\overline{S}(b, c)$ is $K(\sum_{j=1}^{J} H_j)$. On the other hand, mapping of an object $a_i$ also finds $\sum_{j=1}^{J} H_j$ membership degree for both $\underline{R}(a_i, b)$ and $\overline{R}(a_i, b)$. Therefore, with a data set with $I$ objects, the total number of membership degrees of both $\underline{R}(a, b)$ and $\overline{R}(a, b)$ is $I\sum_{j=1}^{J} H_j$.

## 6. Experiment and discussions

### 6.1. Experiment setup

In this section, the BK subproduct is designed as a classifier where three different variants, namely the conventional T1FS-based BK subproduct, IVFS-based BK subproduct and weighted IVFS-based BK subproduct are tested. Three publicly available medical data sets, i.e. Statlog Heart (Heart), Pima Indians Diabetes (Pima) and Wisconsin Diagnosis Breast Cancer (WDBC) are adopted in the experiment. Table 2 is a summary of these data sets. We use 5-fold cross validation (5cv), a widely disseminated approach to conduct the experiment. In this approach, a data set is randomly divided into 5 non-mutual exclusive subsets with approximately equivalent number of instances. Before a subset of the data is tested, the other 4 subsets are used to train the system. Therefore, 5 runs are performed for each data set. For a comparison, we also conduct a reverse modal of this 5cv approach for weighted IVFS-based BK subproduct, i.e. only a subset of the data set (about 20%) is used to train the system, and the remaining 4 subsets (about 80%) are tested.

To prepare the data for training and testing of both IVFS-based BK subproduct and weighted IVFS-based BK subproduct, a set of three standard interval-valued membership functions (Fig. 7) that represent "Low" ($\tilde{F}^*_1$), "Medium" ($\tilde{F}^*_2$) and "High" ($\tilde{F}^*_3$) are defined. These standard membership functions are scaled to the range of the training data of each of the feature, forming membership functions of each feature. Training and fuzzification are performed according to the method described in Section 5.

Same procedure is used for T1FS-based BK subproduct, with the standard interval-valued membership functions replaced with

**Table 2**
A summary of data sets used in the experiment.

| Name | Abbreviation | Instances | Attributes | Classes |
|---|---|---|---|---|
| Statlog Heart | Heart | 270 | 13 | 2 |
| Pima Indians Diabetes* | Pima | 768 | 8 | 2 |
| Wisconsin Diagnostic Breast Cancer | WDBC | 569 | 30 | 2 |

* This data set comes with some missing values.

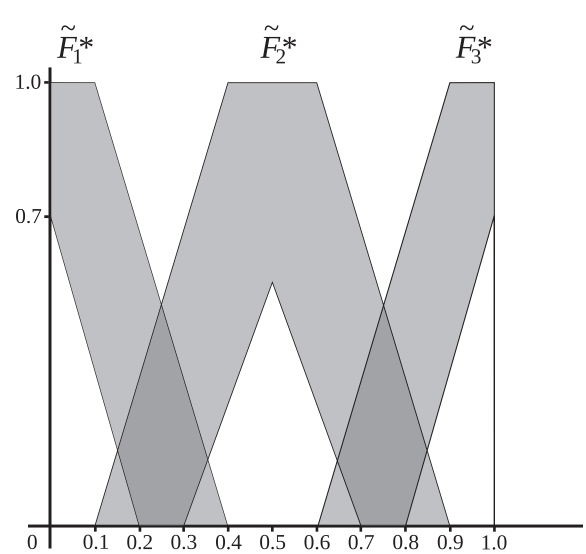


**Fig. 7.** Definition of 3 standard membership functions, $\tilde{F}^*_1, \tilde{F}^*_2$ and $\tilde{F}^*_3$.

**Table 3**
Coordinates of standard membership functions for T1FS.

| | | |
|---|---|---|
| Trapezoidal | $F_1^*$ | $\{(0.0, 0.0), (0.0, 1.0), (0.1, 1.0), (0.4, 0.0)\}$ |
| | $F_2^*$ | $\{(0.1, 0.0), (0.4, 1.0), (0.6, 1.0), (0.9, 0.0)\}$ |
| | $F_3^*$ | $\{(0.6, 0.0), (0.9, 1.0), (1.0, 1.0), (1.0, 0.0)\}$ |
| Triangular | $F_1^*$ | $\{(0.0, 0.0), (0.0, 1.0), (0.4, 0.0)\}$ |
| | $F_2^*$ | $\{(0.1, 0.0), (0.5, 1.0), (0.9, 0.0)\}$ |
| | $F_3^*$ | $\{(0.6, 0.0), (1.0, 1.0), (1.0, 0.0)\}$ |

standard type-1 membership functions. For a fair comparison, a replacement standard type-1 membership function should be: *(i)* "embedded" in the corresponding standard interval-valued membership function, *(ii)* normal and *(iii)* with the same support. Therefore, 2 sets of standard type-1 membership functions are selected (Table 3).

### 6.2. Defuzzification

After the inferences, the results from the inference engines are defuzzified and the following defuzzification strategy is used in the experiment: firstly, center of sets, or centroids (Karnik & Mendel, 1998; Lee, 1990b) are computed from the outputs of the inference engines. Subsequently, compare and rank the centroids of $\tilde{R} \triangleleft \tilde{s}(a_i, c_k)$ for all $k \in K$. Find $c_{k'}$ where $\tilde{R} \triangleleft \tilde{s}(a_i, c_{k'})$ is greater than $\tilde{R} \triangleleft \tilde{s}(a_i, c_k)$ for $k' \in K$ but $k \neq k'$. The ranking shows that $a_i$ has strongest relation to $c_{k'}$. Therefore, $a_i$ belongs to this class.

### 6.3. Weight determination

For the weight determination, features are clustered according to their influence towards the classification results. In this experiment, features are divided into two groups, namely "low" and "high", according to the influence. The $\tilde{F}_1^*$ and $\tilde{F}_3^*$ in Fig. 7 are used as "low" ($\widetilde{W}_1$) and "high" ($\widetilde{W}_2$) weight functions respectively. An algorithm is needed to distinguish whether a feature should belongs to high or low weight group. This weighing algorithm examines each attribute iteratively, using the training data.

Firstly, the data used during the training stage are employed again. Fuzzify the training data by mapping it to the membership function sets $\tilde{F}_{j1}, \tilde{F}_{j2}$ and $\tilde{F}_{j3}$. This form fuzzy relation $\tilde{R}(a^*, b)$ where $a^* \in A^*$ is the training data.

With leaving out a feature $b_j \in B$, perform inferences with IVFS-based BK subproduct, but on training data. Assume $I^*$ is the total number of instances in training data. With this step, for a leaving out feature $b_j$, we find $I^*K$ intervals that measuring the relations of objects and targets without this feature.

With the defuzzification method described in Section 6.2, defuzzify these inference results and compare the defuzzified results with ground truths. Count $G_j$, the number of instances that lead to correct inferences by leaving out $b_j$. The higher the $G_j$, the lower the influence of feature $b_j$ lead to correct inference. Therefore, the weight of $b_j$ is given by $w_j$:

$$w_j = 1 - \frac{G_j}{I^*} \tag{28}$$

The procedure is repeated for all the features to find the corresponding weights. With the results, features can be divided into low and high weight groups.

### 6.4. Results and discussion

This paper aims to improve the BK subproduct with two extensions, namely the IVFSs and a weight parameter. In this section, the improvement on classification accuracy that brings by each extension is discussed separately. The accuracy of a test is the ratio of correctly predicted instances over the complete testing set. The average of all the 5 runs is computed for all the tests. Tables 4–6 present the accuracy of the tests with Heart, Pima and WDBC respectively.

The tables show that our extension on BK subproduct from T1FS to IVFS improves the classification accuracy. Compare to the two implementations of T1FS-based BK subproduct, IVFS-based BK subproduct gives higher accuracy in all cases. The accuracy increments on Heart data set are 1.11% and 1.85%; for Pima data set, the accuracy increments are 2.73% and 3.37%; whereas the reading for WDBC data set is 2.29% and 2.47%. Overall, the improvement rate for the IVFS-based BK subproduct classifier ranged from 1.34% to 5.39%.

Also, it can be noticed that the proposed weight parameter boosted the accuracy of the BK-IVFS. Compare to the non-weighted IVFS-based BK subproduct, the weighted IVFS-based BK subproduct increases the accuracy of Heart, Pima and WDBC datasets by 1.86%, 2.47% and 1.23% respectively. The improvement rate is in the range 1.31–3.35%. The results also show that both the extensions suggested enhance the classification resolution of the BK subproduct. Cumulatively, the improvement on classification accuracy brings by the extensions on the datasets Heart, Pima and WDBC are 3.71%, 6.24% and 3.70% respectively.

To learn more about the IVFS-based BK subproduct based classifier, the results of this study are also compared with seven state-of-the-art works using the same data sets. Among all, Mantas and Abellán (2014) works on all the three data sets as ours, and the others work on two out of the three data sets, i.e. Hu (2013) and Ghaemi and Feizi-Derakhshi (2014) work on Heart and Pima, Pacheco, Alfaro, Casado, Gámez, and García (2012) and Huang and Kechadi (2013) work on Heart and WDBC, and Li and Liu (2010) and Ballings and Van den Poel (2013) work on Pima and WDBC. The comparisons show that the weighted IVFS-based BK subproduct gives the best accuracies in all cases except the WDBC,

**Table 4**
Average accuracy of experiment with Statlog Heart Disease.

| Author | Method | Runs | Train-test ratio | Accuracy (%) |
|---|---|---|---|---|
| Conventional | BK-T1FS (Trapezoidal) | 5 | 4:1 | 82.59 |
| | BK-T1FS (Triangular) | 5 | 4:1 | 81.85 |
| Our proposed | BK-IVFS | 5 | 4:1 | 83.70 |
| | **BK-IVFS (weighted)** | **5** | **4:1** | **85.56** |
| | BK-IVFS (weighted) | 5 | 1:4 | 82.59 |
| Mantas and Abellán (2014) | Credal-C4.5 (no pruning) | 10 | 9:1 | 80.04 |
| | Credal-C4.5 (pruning) | 10 | 9:1 | 80.33 |
| Hu (2013) | RSRC-P | 5 | 4:1 | 84.00 |
| Pacheco et al. (2012) | GRASP | 10 | 9:1 | 78.10 |
| Ghaemi and Feizi-Derakhshi (2014) | FWFOA | 10 | 9:1 | 81.35 |
| Huang and Kechadi (2013) | K-means + FOIL | 5 | 4:1 | 83.80 |

**Table 5**
Average accuracy of experiment with Pima Indian Diabetes.

| Author | Method | Runs | Train-test Ratio | Accuracy (%) |
|---|---|---|---|---|
| Conventional | BK-T1FS (Trapezoidal) | 5 | 4:1 | 70.97 |
| | BK-T1FS (Triangular) | 5 | 4:1 | 69.93 |
| Our proposed | BK-IVFS | 5 | 4:1 | 73.70 |
| | **BK-IVFS (weighted)** | **5** | **4:1** | **76.17** |
| | BK-IVFS (weighted) | 5 | 1:4 | 73.60 |
| Mantas and Abellán (2014) | Credal-C4.5 (no pruning) | 10 | 9:1 | 73.19 |
| | Credal-C4.5 (pruning) | 10 | 9:1 | 74.15 |
| Li and Liu (2010) | SVM Gaussian | 30 | 1:2 | 64.00 |
| | SVM Polynomial | 30 | 1:2 | 62.52 |
| | SVM CPBK | 30 | 1:2 | 71.15 |
| Hu (2013) | RSRC-P | 5 | 4:1 | 74.60 |
| Ghaemi and Feizi-Derakhshi (2014) | FWFOA | 10 | 9:1 | 71.11 |
| Ballings and Van den Poel (2013) | KF | 10 | 1:1 | 71.09–74.48[a] |

[a] In this research, 13 models were developed, and the median of the tests, instead of arithmetic mean were presented.

**Table 6**
Average accuracy of experiment with Wisconsin Diagnosis Breast Cancer.

| Author | Method | Runs | Train-test ratio | Accuracy (%) |
|---|---|---|---|---|
| Conventional | BK-T1FS (Trapezoidal) | 5 | 4:1 | 91.56 |
| | BK-T1FS (Triangular) | 5 | 4:1 | 91.74 |
| Our proposed | BK-IVFS | 5 | 4:1 | 94.03 |
| | **BK-IVFS (weighted)** | **5** | **4:1** | **95.26** |
| | BK-IVFS (weighted) | 5 | 1:4 | 94.03 |
| Mantas and Abellán (2014) | Credal-C4.5 (no pruning) | 10 | 9:1 | 95.08 |
| | Credal-C4.5 (pruning) | 10 | 9:1 | 95.12 |
| Pacheco et al. (2012) | GRASP | 10 | 9:1 | 94.80 |
| Li and Liu (2010) | SVM Gaussian | 30 | 1:4 | 83.14 |
| | SVM Polynomial | 30 | 1:4 | 58.58 |
| | SVM CPBK | 30 | 1:4 | 93.26 |
| Huang and Kechadi (2013) | K-means + FOIL | 5 | 4:1 | 94.60 |
| Ballings and Van den Poel (2013) | KF | 10 | 1:1 | 94.19–95.61[a] |

[a] In this research, 13 models were developed, and the median of the tests, instead of arithmetic mean were presented.

where one of the models proposed by Ballings and Van den Poel (2013) leads with a 0.35% gap.

As we proposed a learning mechanism in the BK subproduct, therefore, as to many other supervised training solutions, the prediction accuracy increases when the number of training data increases. Meanwhile, this method shows an advantage: it manage to train a system with low amount of training data, as long as the pattern of data distribution is obtained. This conclusion can be further affirmed with the experiment results with reverse 5-fold cross validation, where only one subset is used to train the system, whereas the other 4 subsets are test data (train-test ratio 1:4). In the literature, there are very few researches that conducted on low amount of training data, therefore, limited comparisons can be found. Comparing to the state-of-the-art solutions with higher train-test ratio (Tables 4–6), our proposed method that utilized a low amount of training data (ratio 1:4), the obtained accuracies are comparable and in some cases even better.

## 7. Conclusion

The BK subproduct is one of the outstanding reasoning schemes that based on fuzzy relational theory. In this paper, we study this reasoning scheme as a fuzzy inference engine. Furthermore, the fuzzy inference engine is implemented as a classifier to classify three widely adopted public data sets. Experiment results show that this inference engine outperform other classifiers that work on the same datasets.

While many others still working on fuzzy inference systems that based on rules, this research continues the exploration of performing inferences when rules can not be defined or well defined. The contributions of this research are threefold: (i) A subsethood measure for IVFS is defined, where we extend the BK subproduct from T1FS-based to IVFS-based. This extension increases the ability of the BK subproduct to capture uncertainty in reasoning. (ii) Incorporating the weight parameter in the BK subproduct reasoning, with theoretical explanation on the validity of this weight parameter. (iii) A novel training mechanism is developed to construct the knowledge based for the BK subproduct based inference engines. This mechanism solved the problem where expert knowledge is required to construct systems that based on BK subproduct.

In term of theoretical implication, this research proposed a high reliability inference engine that is rule-free, tolerate to uncertainty with IVFS and consider features weight. Compare to many other fuzzy inference engines which defining rules is compulsory, this inference engine works without rules. This makes the development of inference engines on systems which rules are ill-define become easy (Kolodner, 1992). Furthermore, a new automatic learning mechanism is also proposed in this research. This automatic learning mechanism not only fast and efficient, but also able to train a system with limited training data. In the demonstration in Section 6, we show that even only 20% of the data is used in the training, the experiment results still outperform some other state-of-the-art researches.

The strength of this BK subproduct-based inference engine includes high accuracy, strong mathematical background (Stepnicka & De Baets, 2010, 2013), tolerate to uncertainties and etc. However, this inference engine associates with a weakness in the weight determination. In case of a dataset with $I^*$ training data, $J$ features and $K$ possible targets, a number of $I^*JK$ inferences in the form of Eq. (10) are needed to determine the weights associate to all the features. This weight computation procedure is relatively high computational cost.

Therefore, one of the future research direction is on a better weight determination algorithm. This algorithm is a key to boost the efficiency and accuracy of the weighted BK subproduct based inference engines. The advancement of feature selection algorithm Luukka (2011) may provide clues in developing this mechanism. The second future research direction is on the fine tuning of membership functions. Three standard membership functions are used across all the features in all the data sets. It is reasonable to believe that with an algorithm to fine tune of the standard membership functions (Fig. 7) and weight functions, the classification accuracy can be improve. Technology such as artificial neural network or similar can be considered in this work in the future. Last but not least, the developing the application of BK subproduct is another future research direction. The BK subproduct is a study on the composition of relations between sets that are not directly related. Classifier is only one of the possible applications of BK subproduct. Since relations provides important notions in human reasoning, it is possible to apply BK subproduct in other problems such as data mining, computing with words and engineering control.

## Acknowledgement

This research is supported by the High Impact MoE Grant UM.C/625/1/HIR/MoE/FCSIT/08, H-22001-00-B00008 from the Ministry of Education Malaysia.

## References


Abdullah, L., & Najib, L. (2014). A new type-2 fuzzy set of linguistic variables for the fuzzy analytic hierarchy process. *Expert Systems with Applications, 41*(7), 3297–3305.

Ait Laasri, E. H., Akhouayri, E.-S., Agliz, D., Zonta, D., & Atmani, A. (2015). A fuzzy expert system for automatic seismic signal classification. *Expert Systems with Applications, 42*(3), 1013–1027.

Ballings, M., & Van den Poel, D. (2013). Kernel Factory: An ensemble of kernel machines. *Expert Systems with Applications, 40*(8), 2904–2913.

Bodenhofer, U., Dankova, M., Stepnicka, M., & Novak, V. (2007). A plea for the usefulness of the deductive interpretation of fuzzy rules in engineering applications. *IEEE International Conference on Fuzzy Systems*, 1–6.

Bugarski, V., Bačkalić, T., & Kuzmanov, U. (2013). Fuzzy decision support system for ship lock control. *Expert Systems with Applications, 40*(10), 3953–3960.

Bui, L.-d., & Kim, Y.-g. (2006). An obstacle-avoidance technique for autonomous underwater vehicles based on BK-products of fuzzy relation. *Fuzzy Sets and Systems, 157*, 560–577.

Camastra, F., Ciaramella, A., Giovannelli, V., Lener, M., Rastelli, V., Staiano, A., et al. (2015). A fuzzy decision system for genetically modified plant environmental risk assessment using mamdani inference. *Expert Systems with Applications, 42*(3), 1710–1716.

Civanlar, M. R., & Trussell, H. J. (1986). Constructing membership functions using statistical data. *Fuzzy Sets and Systems, 18*(1), 1–13.

D'Andrea, E., & Lazzerini, B. (2013). A hierarchical approach to multi-class fuzzy classifiers. *Expert Systems with Applications, 40*(9), 3828–3840.

Dong, W. M., & Wong, F. S. (1987). Fuzzy weighted averages and implementation of the extension principle. *Fuzzy Sets and Systems, 21*, 183–199.

Egrioglu, E., Aladag, C. H., & Yolcu, U. (2013). Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks. *Expert Systems with Applications, 40*(3), 854–857.

Fazel Zarandi, M., Rezaee, B., Turksen, I., & Neshat, E. (2009). A type-2 fuzzy rule-based expert system model for stock price analysis. *Expert Systems with Applications, 36*(1), 139–154.

Ghaemi, M., & Feizi-Derakhshi, M.-R. (2014). Forest optimization algorithm. *Expert Systems with Applications, 41*(15), 6676–6687.

Groenemans, R., Ranst, V. E., & Kerre, E. E. (1997). Fuzzy relational calculus in land evaluation. *Geoderma, 77*, 283–298.

Gupta, Y., Saini, A., & Saxena, A. (2014). A new fuzzy logic based ranking function for efficient information retrieval system. *Expert Systems with Applications*.

Hu, Y.-C. (2013). Rough sets for pattern classification using pairwise-comparison-based tables. *Applied Mathematical Modelling, 37*(12-13), 7330–7337.

Huang, Y., & Kechadi, T. (2013). An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications, 40*(14), 5635–5647.

Ishibuchi, H., & Yamamoto, T. (2005). Rule weight specification in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems, 13*(4), 428–435.

Karnik, N., & Mendel, J.M. (1998). Type-2 fuzzy logic systems: Type-reduction. In *IEEE international conference on systems, man, and cybernetics* (pp. 2046–2051).

Klir, G. J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications*. New Jersey: Prentice Hall.

Kohout, L. J. (2009). Power sets, implications and set inclusions revisited – retrospect and prospect: A review of Bandler and Kohout's paper and a survey of 30 years of subsequent developments. *Views on fuzzy sets and systems from different perspectives* (Vol. 243, pp. 129–159). Springer.

Kohout, L. J., & Bandler, W. (1980a). Fuzzy power sets and fuzzy implication operators. *Fuzzy Sets and Systems, 4*, 13–30.

Kohout, L. J., & Bandler, W. (1980b). Semantics of implication operators and fuzzy relational products. *International Journal of Man–Machine Studies, 12*(1), 89–116.

Kohout, L. J., & Bandler, W. (1985). Relational-product architectures for information processing. *Information Sciences, 37*, 25–37.

Kohout, L. J., & Bandler, W. (1992). How the checklist paradigm elucidates the semantics of fuzzy inference. *IEEE International Conference on Fuzzy Systems*, 571–578.

Kohout, L. J., Stabile, I., Kalantar, H., & San-Andres, F. (1995). Parallel interval-based reasoning in medical knowledge-based system Clinaid. *Reliable Computing, 1*(2), 109–140.

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review, 6*(1), 3–34.

Lam, H., & Seneviratne, L. D. (2008). Stability analysis of interval type-2 fuzzy-model-based control systems. *IEEE Systems, Man, and Cybernetics, Part B, 38*(3), 617–628.

Lee, C. C. (1990a). Fuzzy logic in control systems: Fuzzy logic controller. Part I. *IEEE Transactions on Systems, Man, and Cybernetics, 20*(2), 404–418.

Lee, C. C. (1990b). Fuzzy logic in control systems: Fuzzy logic controller, part II. *IEEE Transactions on Systems, Man, and Cybernetics, 20*(2), 419–435.

Li, D.-C., & Liu, C.-W. (2010). A class possibility based kernel to increase classification accuracy for small data sets using support vector machines. *Expert Systems with Applications, 37*(4), 3104–3110.

Lim, C.K., & Chan, C.S. (2012). Fuzzy set and multi descriptions property. In *IEEE international conference on fuzzy systems* (pp. 1–8).

Lim, C. K., & Chan, C. S. (2011). Logical connectives and operativeness of BK subtriangle product in fuzzy inferencing. *International Journal of Fuzzy Systems, 13*(4), 237–245.

Liu, W., Han, J., & Lu, X. (2013). A high speed railway control system based on the fuzzy control method. *Expert Systems with Applications, 40*(15), 6115–6124.

Liu, F., & Mendel, J. M. (2008). Aggregation using the fuzzy weighted average as computed by the Karnik–Mendel algorithms. *IEEE Transactions on Fuzzy Systems, 16*(1), 1–12.

Liu, P., Yang, W.-T., & Yang, C.-E. (2013). Robust observer-based output feedback control for fuzzy descriptor systems. *Expert Systems with Applications, 40*(11), 4503–4510.

Luukka, P. (2011). Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications, 38*(4), 4600–4607.

Mandal, S., & Jayaram, B. (2013). Monotonicity of SISO fuzzy relational inference mechanism with Yager's Class of fuzzy implications. *Pattern Recognition and Machine Intelligence*, 569–574.

Mantas, C. J., & Abellán, J. (2014). Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data. *Expert Systems with Applications, 41*(10), 4625–4637.

Melin, P., Olivas, F., Castillo, O., Valdez, F., Soria, J., & Valdez, M. (2013). Optimal design of fuzzy classification systems using PSO with dynamic parameter adaptation through fuzzy logic. *Expert Systems with Applications, 40*(8), 3196–3206.

Mendel, J. M. (2000). Interval type-2 fuzzy logic systems: Theory and design. *IEEE Transactions on Fuzzy Systems, 8*(5), 535–550.

Mendel, J. M. (2003). Type-2 fuzzy sets: Some questions and answers. *IEEE Neural Networks Society*, 10–13.

Mendel, J. M. (2009). Enhanced Karnik–Mendel algorithms. *IEEE Transactions on Fuzzy Systems, 17*(4), 923–934.

Mendel, J. M., & John, R. (2002). Type-2 fuzzy sets made simple. *IEEE Transactions on Fuzzy Systems, 10*(2), 117–127.

Mizumoto, M., & Zimmermann, H. (1982). Comparison of fuzzy reasoning methods. *Fuzzy Sets and Systems, 8*(3), 253–283.

Nguyen, H., & Kreinovich, V. (2008). Computing degrees of subsethood and similarity for interval-valued fuzzy sets: Fast algorithms. In *Proceedings of the 9th international conference on intelligent technologies* (pp. 47–55).

Nguyen, H.-T., Dawal, S. Z. M., Nukman, Y., & Aoyama, H. (2014). A hybrid approach for fuzzy multi-attribute decision making in machine tool selection with consideration of the interactions of attributes. *Expert Systems with Applications, 41*(6), 3078–3090.

Pacheco, J., Alfaro, E., Casado, S., Gámez, M., & García, N. (2012). A GRASP method for building classification trees. *Expert Systems with Applications, 39*(3), 3241–3248.

Petrović, D. V., Tanasijević, M., Milić, V., Lilić, N., Stojadinović, S., & Svrkota, I. (2014). Risk assessment model of mining equipment failure based on fuzzy logic. *Expert Systems with Applications, 41*(18), 8157–8164.

Rickard, J. T., Aisbett, J., & Greb, G. (2009). Fuzzy subsethood for fuzzy sets of type-2 and generalized type-n. *IEEE Transactions on Fuzzy Systems, 17*(1), 50–60.

Roh, S.-B., Pedrycz, W., & Ahn, T.-C. (2014). A design of granular fuzzy classifier. *Expert Systems with Applications*.

Samantra, C., Datta, S., & Mahapatra, S. S. (2014). Risk assessment in it outsourcing using fuzzy decision-making approach: An indian perspective. *Expert Systems with Applications, 41*(8), 4010–4022.

Samuel, O., Omisore, M., & Ojokoh, B. (2013). A web based decision support system driven by fuzzy logic for the diagnosis of typhoid fever. *Expert Systems with Applications, 40*(10), 4164–4171.

Seki, H., & Mizumoto, M. (2011). On the equivalence conditions of fuzzy inference methods – Part 1: Basic concept and definition. *IEEE Transactions on Fuzzy Systems, 19*(6), 1097–1106.

Son, L. H. (2015). DPFCM: A novel distributed picture fuzzy clustering method on picture fuzzy sets. *Expert Systems with Applications, 42*(1), 51–66.

Stepnicka, M., & De Baets, B. (2010). Monotonicity of implicative fuzzy models. In *Proceedings of the IEEE international conference on fuzzy systems* (pp. 1–7).

Stepnicka, M., & De Baets, B. (2013). Implication-based models of monotone fuzzy rule bases. *Fuzzy Sets and Systems, 232*, 134–155.

Štěpnička, M., & Holčapek, M. (2014). Fuzzy relational compositions based on generalized quantifiers. *Information Processing and Management of Uncertainty in Knowledge-Based Systems, 443*, 224–233.

Stepnicka, M., & Jayaram, B. (2010). On the suitability of the Bandler–Kohout subproduct as an inference mechanism. *IEEE Transactions on Fuzzy Systems, 18*(2), 285–298.

Suparta, W., & Alhasa, K. M. (2014). Modeling of zenith path delay over antarctica using an adaptive neuro fuzzy inference system technique. *Expert Systems with Applications*.

Vats, E., Lim, C.K., & Chan, C.S. (2012). A BK subproduct approach for scene classification. In *Proceedings of the IIEEJ image electronics and visual computing workshop* (pp. 1–5). Kuching, Malaysia.

Wu, D., & Mendel, J. M. (2007). Aggregation using the linguistic weighted average and interval type-2 fuzzy sets. *IEEE Transactions on Fuzzy Systems, 15*(6), 1145–1161.

Wu, D., & Mendel, J. M. (2008). Corrections to aggregation using the linguistic weighted average and interval type-2 fuzzy sets. *IEEE Transactions on Fuzzy Systems, 16*(6), 1664–1666.

Xing, H.-J., & Ha, M.-H. (2014). Further improvements in feature-weighted fuzzy c-means. *Information Sciences, 267*, 1–15.

Yager, R. R. (1980). An approach to inference in approximate reasoning. *International Journal of Man–Machine Studies, 13*(3), 323–338.

Yang, M.-S., & Lin, D.-C. (2009). On similarity and inclusion measures between type-2 fuzzy sets with an application to clustering. *Computers & Mathematics with Applications, 57*(6), 896–907.

Yew, K. M., & Kohout, L. J. (1997). Interval-based reasoning in medical diagnosis. In *Proceedings intelligent information systems* (pp. 32–36). IEEE Comput. Soc..

Yuste, A. J., Triviño, A., & Casilari, E. (2013). Type-2 fuzzy decision support system to optimise manet integration into infrastructure-based wireless systems. *Expert Systems with Applications, 40*(7), 2552–2567.

Zadeh, L.A. (1975). Calculus of fuzzy restrictions. Electronics research laboratory, University of California.

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

Zheng, G., Xiao, J., Zhang, Y., & Shi, G. (2010). An inclusion measure between general type-2 fuzzy sets. In *International conference on fuzzy systems and knowledge discovery* (pp. 99–103).