Evidential K-NN Classification with Enhanced Performance via Optimizing A Class of Parametric Conjunctive t-Rules☆

Zhi-gang Su[1], Thierry Denoeux[2], Yong-sheng Hao[1], Ming Zhao[3]

[1] School of Energy and Environment, Southeast University, Nanjing, Jiangsu 210096, China

[2] Sorbonne Universités, Université de Technologie de Compiègne, CNRS, Heudiasyc, Compiègne, France

[3] Research Institute of Yunnan Power Grid Co. Ltd., Kunming, Yunnan, China

Abstract:

Dempster's rule of combination is commonly used to pool distinct/independent bodies of evidence in the evidential k-nearest neighbor (K-NN) classifier, which sometimes limits the performance of this classifier. To solve this problem, we propose a class of parametric conjunctive combination rules based on a new family of triangular norms with selectable functions and tunable parameters. We show that the performance of the evidential K-NN classifier can be enhanced via this class of so-called parametric conjunctive t-rules when appropriate functions and parameters are selected. Numerical simulations validate our conclusions.

Keywords: Evidence theory; Dempster-Shafer theory; belief functions; combination rules; pattern recognition; supervised learning

1. Introduction

The last four decades have seen the emergence and development of the theory of belief functions, which has gained increasing interest as a conceptual framework for modeling partial knowledge and reasoning under uncertainty, see [4-7, 26-27, 32-35, 38-39] for example, and the literature therein. Among the applications of this theory, the evidential K-NN classifier has been successfully used in a wide range of classification problems [5, 11, 16-20, 34-36, 39].

In the evidential K-NN classifier, Dempster's rule is commonly used to pool bodies of evidence, due to its simplicity. This rule implicitly requires that the bodies of evidence to be combined are distinct [28-29]. In other words, the information sources should be independent. However, it is difficult to guarantee the independence of information sources in many practical applications. How to enhance the performances of the evidential K-NN classifier via optimizing a class of possibly parametric combination rules to account for dependence of the sources is still an open question.

To relax the distinctness/independence assumption, some alternative combination rules have been proposed, see, for instance [1-3, 6, 9, 13 21-25, 29-30, 33]. (Note that we do not aim to review the literature on combination rules.) Among these alternatives, the *cautious* rule [6] is a conjunctive rule based on the least commitment principle (LCP) [30]. The LCP stipulates that one should never give more beliefs than justified by the available information; it thus promotes a cautious attitude. The cautious rule can be directly derived from the conjunctive weight function [31], an equivalent representation of a non-dogmatic belief function. After providing a new justification of conjunctive rule in [22], Pichon and Denoeux generalized it and proposed the triangular-norm (t-norm for short) based conjunctive rules, called

conjunctive t-rule [22-24]. They showed that the cautious rule can be seen as a particular conjunctive t-rule when choosing a particular t-norm defined on (0, +∞], whereas the conjunctive rule belongs to a family of rules defined by selecting a suitable uninorms on (0, +∞] having 1 as neutral element.

As remarked in [8], having only one rule is "not very fortunate in the scope of Artificial Intelligent (AI), where one tries to simulate the human mind rather than force behavior rules issued from formal arguments". Indeed, the descriptive nature of AI calls for more flexibility in the choice of combination rules, provided that those other rules satisfy at least a few basic and reasonable requirements. From this viewpoint, the conjunctive t-rules or any other alternatives should be investigated in depth so as to obtain some new combination rules. In fact, the conjunctive t-rules were formulated using a particular family of t-norms defined on (0, +∞]. Whereas the current conjunctive t-rules seem to have sufficient ability to account for the dependence of bodies of evidence, it will be shown in this paper that some more effort is needed so that these rules can be used to enhance the performance of the evidential K-NN classifier.

Motivated by the above considerations, this paper aims to enhance the performance of the evidential K-NN classifier by optimizing a class of parametric combination rules including conjunctive t-rules, called *parametric conjunctive t-rules* (PCTR). More precisely, a scheme is first proposed to construct a t-norm with selectable functions and tunable parameters. With the help of this new t-norm, a class of parametric conjunctive t-rules is then proposed and applied to pool bodies of evidence in the evidential K-NN classifier. We will show that the proposed parametric conjunctive t-rules can enhance the performance of the evidential K-NN classifier if appropriate functions and parameters are selected.

The rest of this paper is organized as follows. The theory of belief function is first briefly recalled in Section 2. The parametric conjunctive t-rules are proposed after constructing a class of t-norms in Section 3. In Section 4 we conduct some numerical experiments to study the performances of the proposed combination rules and the evidential K-NN classifier. The last section concludes the paper.

2. Preliminaries: Background on belief functions

In this section, we briefly recall some notions of the theory of belief functions [4, 26, 32] needed in the rest of the paper. Given a frame of discernment $\Omega = \{\omega_1, \omega_2, ..., \omega_c\}$, a *mass function* or *basic belief assignment* (BBA) m is defined as a mapping from 2^Ω to [0, 1] such that

$$\sum_{A \subseteq \Omega} m(A) = 1 , \tag{1}$$

where the subsets A of Ω such that $m(A) > 0$ are called the focal sets of m.

A BBA is said to be *categorical* if it has only one focal set. In particular, the vacuous BBA verifies $m(\Omega) = 1$: it corresponds to complete ignorance. A BBA is said to be non-dogmatic if Ω is a focal set. A BBA m is *subnormal* if the empty set ϕ is a focal set, and *normal* otherwise. A subnormal BBA m such that $m(\phi) < 1$ can be transformed into a normal BBA m^* by the *normalization* operation defined as:

$$m^*(A) = \begin{cases} \dfrac{m(A)}{1-m(\phi)}, & if \ \ \phi \neq A \subseteq \Omega, \\ 0, & otherwise. \end{cases} \qquad (2)$$

According to Shafer [26], function m is said to be a *simple* BBA (SBBA) if it has the following form: $m(A) = 1-w$, $m(\Omega) = w$ for some $A \subset \Omega$ and some $w \in [0, 1]$. Smets [31] defined a generalized SBBA (GSBBA) as a function $\mu: 2^{\Omega} \rightarrow (0, +\infty)$ such that

$$\begin{cases} \mu(A) = 1-w, \\ \mu(\Omega) = w, \\ \mu(B) = 0, & \forall B \in 2^{\Omega} \setminus \{A, \Omega\}, \end{cases} \qquad (3)$$

for some $A \neq \Omega$ and $w \in (0, +\infty)$. A SBBA or GSBBA can be denoted by A^w for simplicity. Smets [31] showed that any non-dogmatic BBA m admits a canonical decomposition as the combination of GSBBAs $A^{w(A)}$ focused on each of the strict subsets A of Ω with weights $w(A)$. Function $w: 2^{\Omega} \setminus \{\Omega\} \rightarrow (0, +\infty)$ is called the *conjunctive weight function* associated to m. A BBA is said to be *separable* if it is canonically decomposed into SBBAs, i.e., if $w(A) \leq 1$ for all A.

There are other equivalent representations of a BBA such as the *belief*, *plausibility*, and *commonality* functions. In particular, the commonality function is defined as

$$q(A) = \sum_{A \subseteq B} m(B), A \subseteq \Omega . \qquad (4)$$

The weights $w(A)$ in Eq. (3) can be obtained for every $A \subset \Omega$ according to

$$w(A) = \prod_{A \subseteq B} q(B)^{(-1)^{|B|-|A|+1}} . \qquad (5)$$

The combination of bodies of BBAs plays a critical role in the theory of belief functions. Let m_1 and m_2 be two BBAs. The *conjunctive combination* of m_1 and m_2, denoted by $\bigcirc\!\!\!\cap$, results in a new BBA defined as

$$m_{1 \bigcirc\!\!\!\cap 2}(A) = \sum_{B \cap C = A} m_1(B) m_2(C), A \subseteq \Omega. \qquad (6)$$

The conjunctive rule allows some mass to be assigned to the empty set. Dempster's rule, denoted by $\oplus$, consists in the conjunctive rule followed by normalization (2). Both rules are commutative, associative and admit the vacuous BBA as a unique neutral element. The conjunctive rule can be expressed using conjunctive weight functions as

$$m_{1 \bigcirc\!\!\!\cap 2}(A) = \bigcirc\!\!\!\cap_{A \subset \Omega} A^{w_1(A) \times w_2(A)} , \qquad (7)$$

where $\times$ denotes the product. By using the minimum, denoted by $\wedge$, instead of the product, the cautious rule can be derived for two non-dogmatic BBAs as

$$m_{1 \bigotimes 2}(A) = \bigcirc\!\!\!\cap_{A \subset \Omega} A^{w_1(A) \wedge w_2(A)} . \qquad (8)$$

Pichon and Denoeux provided a new justification of the conjunctive and cautious combination rules in [24], and proposed the so-called *conjunctive t-rules* by replacing the minimum $\wedge$ by another t-norm on $(0, +\infty]$ such as [22, 23]

$$x \, \mathcal{T}_{\top\top'} \, y = \begin{cases} x \top y, & x \vee y \leq 1, \\ \left(1 - \left(1 - \dfrac{1}{x} \right) \top' \left(1 - \dfrac{1}{y} \right) \right)^{-1}, & x \wedge y > 1, \\ x \wedge y, & \textit{otherwise}, \end{cases} \qquad (9)$$

for all x, y in $(0, +\infty]$, where $\vee$ denotes the maximum operator, and $\top$ and $\top'$ are t-norms on $[0, 1]$ [14].

Finally, the *pignistic* transformation [32] can be used to transform a combined BBA into a probability measure in order to make decision. It is defined as

$$BetP_m \left(\left\{ \omega_q \right\} \right) = \sum_{\omega_q \in A \subseteq \Omega} \frac{m^*(A)}{|A|}, \qquad (10)$$

for all ω in Ω.

3. Main results: Rules and Method

After introducing a family of t-norms and construct methods, a class of parametric conjunctive t-rules will first be proposed from the conjunctive t-rules in Section 3.1. The evidential K-NN classifier based on parametric conjunctive t-rules will then be presented in Section 3.2.

3.1 Parametric conjunctive t-rules.

Definition 1 [23] A t-norm on $(0, +\infty]$ is a binary operation on $(0, +\infty]$ that is commutative, associative, monotonic, and admits $+\infty$ as neutral element.

Generalizing Eq. (9), the following proposition provides a scheme to construct a class of t-norm on $(0, +\infty]$ from t-norms on $[0, 1]$ as well as some selectable functions.

Proposition 1 Consider an increasing function $g_1(x) : [0, 1] \rightarrow [0, 1]$ such that $g_1(0) = 0$, $g_1(1) = 1$, and a strictly decreasing function $g_2(x) : [1, +\infty] \rightarrow [0, 1]$ such that $g_2(1) = 1$, $g_2(+\infty) = \lim_{x \to +\infty} g_2(x) = 0$. Let $g_2^{-1}(x)$ be the inverse function of $g_2(x)$. Then, the binary operation defined by the following equation is a t-norm on $(0, +\infty]$

$$x \, \mathcal{T} \, \overset{g_i}{_{\top\top'}} \, y = \begin{cases} g_1(x) \top g_1(y), & x \vee y \leq 1, \\ g_2^{-1} \left(N \left(N \left(g_2(x) \right) \top' N \left(g_2(y) \right) \right) \right), & x \wedge y > 1, \\ x \wedge y, & \textit{otherwise}, \end{cases} \qquad (11)$$

where $\top$ and $\top'$ are classical positive t-norms on $[0, 1]$ such that $x \top y > 0$, $\forall \, x > 0$, $y > 0$, and $N(x)$ is the negation operation on $[0, 1]$ such that $N(x) = 1 - x$.

Proof: See in Appendix.

Remark 1 Proposition 1 provides a scheme to construct t-norms on $(0, +\infty]$. By choosing different functions $g_i(x)$ and t-norms $\top$ and $\top'$ on $[0, 1]$, infinite families of t-norms $(0, +\infty]$ can be obtained. The t-norms $\top$ and $\top'$ can be selected, e.g., as the *minimum, product, Dubois-Prade,* and *Frank* t-norms defined, respectively, as

$$T_M(x, y) = \min(x, y) \,,$$

$$T_P(x, y) = xy \,,$$

$$T_{DP}(x,y) = \frac{xy}{\max\{x,y,\lambda\}}, \lambda \in [0,1],$$

and

$$T_F(x,y) = \log_\lambda\left(1 + \frac{(\lambda^x - 1)(\lambda^y - 1)}{\lambda - 1}\right)$$

with $\lambda > 0$ and $\lambda \neq 1$. If $\top = \top' = \wedge$, then we recover the minimum. It is evident that Eq. (9) is a special case of the t-norms (11) with $g_1(x) = x$ and $g_2(x) = x^{-1}$. There are many choices of functions $g_i(x)$, $i = 1, 2$. In this paper, we use the power function $g_1(x) = x^\beta$ with positive constant β, and exponential or power functions $g_2(x)$ such that $g_2(x) = e^{\chi(1-x)}$ or $g_2(x) = x^{-\chi}$ with positive constant χ. Fig. 1 shows the curves of $g_i(x)$ in some cases.
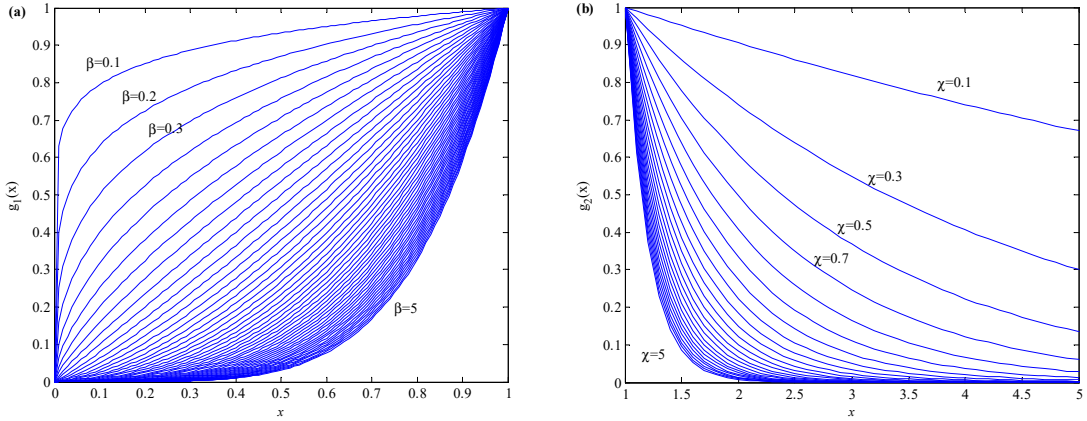

Fig. 1 Curves of functions (a) $g_1(x) = x^\beta$ and (b) $g_2(x) = e^{\chi(1-x)}$

Remark 2 Note that, as will be seen, all BBAs to be combined in the evidential K-NN classification setting are obtained by discounting categorical BBAs and are, therefore, separable. Hence, the t-norm (11) can be simplified as follows

$$x \; \mathcal{T}_\top^{g_1} \; y = g_1(x) \top g_1(y), \tag{12}$$

for all x, y in (0, 1], and t-norm $\top$ on [0, 1].

In what follows, we focus on special case (12), although some definitions and propositions are still introduced based on Proposition 1, which is interesting for its own sake. The following example illustrates some t-norms constructed using Eq. (12).

Example 1

Suppose $g_1(x) = x^\beta$ and $\top = T_{DP}$. We can have the following t-norm:

$$x \; \mathcal{T}_\top^{g_1} \; y = T_{DP}(x^\beta, y^\beta), \quad x, y \in [0,1],$$

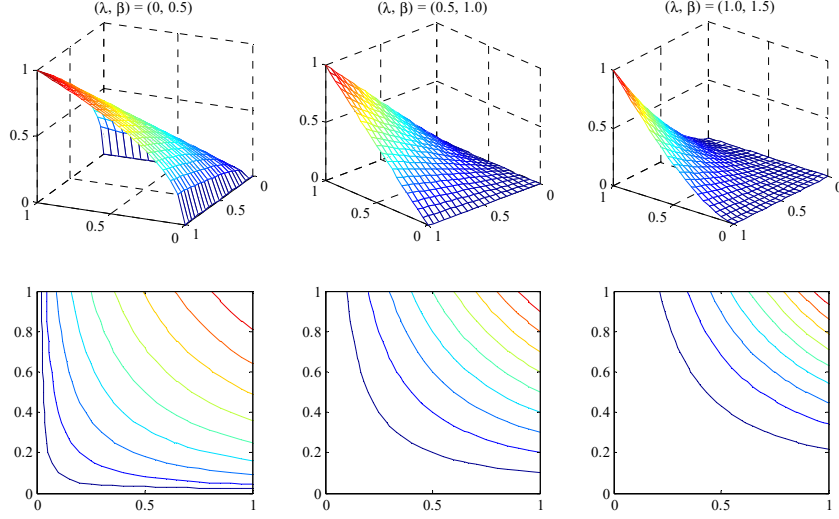whose surfaces as well as contours are shown in Fig. 2.

Fig. 2 Surfaces (top) and contours (bottom) of t-norms in Example 1

In the following, we will simplify the notation $\mathcal{T}_{\top\top'}^{g_i}$ and $\mathcal{T}_{\top}^{g_1}$ as $\mathcal{T}$ whenever there is no risk of confusion. With the help of Proposition 1, we are ready to present the parametric conjunctive t-rules.

Proposition 2 Let $\mathcal{T}$ be the t-norm defined in Proposition 1, and let w_1 and w_2 be the conjunctive weight functions associated to any two non-dogmatic BBAs m_1 and m_2. Then, the weight function defined by

$$w_{1\,\textcircled{\tiny T}_w 2}(A) = w_1(A)\,\mathcal{T}\,w_2(A), \quad \forall A \subset \Omega,$$

is a conjunctive weight function associated to some non-dogmatic BBA $m_{1\,\textcircled{\tiny ?}_w 2}$.

Proof: See in Appendix.

Definition 2 (*Parametric conjunctive t-rules, PCTR*) Let $\mathcal{T}$ be the t-norm in Proposition 1, and let w_1 and w_2 be the conjunctive weight functions associated to any two non-dogmatic BBAs m_1 and m_2. The parametric conjunctive combination of m_1 and m_2 can be calculated according to

$$m_{1\,\textcircled{\tiny ?}_w 2}(A) = m_1 \,\textcircled{\tiny ?}_w^{\,g_i,\top,\top'}\, m_2(A) = \bigcirc_{A\subset\Omega} A^{w_1(A)\,\mathcal{T}\,w_2(A)}. \tag{13}$$

Remark 3 The t-norm based combination rules in Definition 2 are called *parametric conjunctive t-rules*. They generalize the conventional conjunctive t-rules by introducing selectable functions $g_i(x)$ with tunable parameters. When $g_1(x) = x$ and $g_2(x) = x^{-1}$, the parametric conjunctive t-rules degenerate to the conventional conjunctive t-rules. As will be seen in Section 4, the selectable functions $g_i(x)$ provide more flexibly to deal with non-distinct bodies of evidence.

Remark 4 The parametric conjunctive t-rules have two limitations. First, they do not have the vacuous BBA as neutral element. Second, the BBAs to be combined need to be non-dogmatic as only non-dogmatic BBAs have a unique conjunctive canonical decomposition. In practice, however, it may be argued that most states of belief, being based on imperfect and not entirely conclusive evidence, should be represented by non-dogmatic BBAs even if the mass assigned to Ω is very small.

3.2. Evidential K-NN classifier via PCTR

We consider the problem of classifying patterns into c classes. The set of classes is denoted by $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$. The available information is assumed to consist of a training set $TR = \{(x_i, m_i) \mid i = 1, 2, \ldots, n\}$ with d-dimensional patterns x_i and their corresponding class labels represented by mass functions $m_i(\{\omega_q\}) = 1$.

Let x be a new vector to be classified on the basis of the information contained in TR. Each pair (x_i, m_i) constitutes a distinct body of evidence regarding the class membership of x. If x is "close" to x_i, according to the relevant metric $d(., .)$ (e.g., the Euclidean metric), we will be inclined to believe that both vectors belong to the same class. On the contrary, if $d(x, x_i)$ is very large, the consideration of x_i will leave us in a situation of almost complete ignorance concerning the class of x. Consequently, this body of evidence may be postulated to induce a BBA $m[x|x_i](\bullet)$ over Ω defined by

$$m\left[x|x_i\right](A) = \begin{cases} \alpha\phi_q\left(d_i\left(x, x_i\right)\right), & A = \{\omega_q\}, \\ 1 - \alpha\phi_q\left(d_i\left(x, x_i\right)\right), & A = \Omega, \\ 0, & otherwise. \end{cases} \tag{14}$$

where α is a parameter such that $0 < \alpha < 1$, and ϕ_q is a decreasing function verifying

$$\phi_q(0) = 1, \text{ and } \lim_{d_i \to \infty} \phi_q(d_i) = 0. \tag{15}$$

A popular choice for decreasing function ϕ_q is [5]

$$\phi_q(d_i) = \exp\left(-\gamma_q d_i^2\right), \tag{16}$$

where γ_q is a positive parameter associated to class ω_q.

As a result of considering each training pattern in turn, we obtain n BBAs that can be combined by using the parametric conjunctive t-rules to synthesize our final belief regarding the class of x.

$$m\left[x|TR\right] = m\left[x|x_1\right] \underset{w}{\textcircled{T}}^{g_i, \top, \top'} \cdots \underset{w}{\textcircled{T}}^{g_i, \top, \top'} m\left[x|x_n\right]. \tag{17}$$

Since training patterns situated far from x actually provide very little information, it is sufficient to consider the k-nearest neighbors of x in TR. Furthermore, when $m[x|x_i]$ is simple, we may obtain the following simple expression for $m[x|x_i]$:

$$m\left[x|x_i\right] = \{\omega_q\}^{1 - \alpha\phi_q(d_i(x, x_i))} \tag{18}$$

Accordingly, (17) becomes to

$$m\left[x|S_K\right] = \underset{q=1}{\overset{c}{\textcircled{\bigcirc}}} \{\omega_q\}^{\top_{\{i \in S_K(x)\}} g_1\left(1 - \alpha\phi_q(d_i(x, x_i))\right)}, \tag{19}$$

where $S_K(x)$ contains the indexes of the k-nearest neighbors of x in TR.

Finally, the predicted label of pattern x can be obtained as

$$\omega(x) = \underset{1 \leq q \leq c}{\arg\max} \, BetP_{m[x|S_K]}\left(\{\omega_q\}\right). \tag{20}$$

Remark 5 Due to the inequality $g_1(1 - \alpha\phi_q(d_i(x, x_i))) \leq 1$, BBAs $m[x|x_i]$ are separable. Hence, the parametric conjunctive

t-rules depend neither on the t-norm $\top'$ nor on function $g_2(x)$ in the classification setting, as stated in Remark 2. In the performance analysis of evidential K-NN classifier, we will focus on the application of (12) to optimize parameters in function $g_1(x)$ and the t-norm. In this case, only parameter in the t-norm needs to be optimized to enhance the performance of the evidential K-NN classifier when using the conventional conjunctive t-rule (i.e., PCTR with $g_1(x) = x$, see PCTR-λ in Table 3). This is the main difference between our parametric conjunctive t-rules and the conventional conjunctive t-rules.

4. Simulations

This section consists of two parts. In Section 4.1, a numerical example is used to show that the true parametric conjunctive t-rule can be recovered by minimizing a distance function. In Section 4.2, we assess the performance of the evidential K-NN classifier with the parametric conjunctive t-rules.

4.1 Approximation of a combination rule by a parametric conjunctive t-rule

This example shows how to find an optimal parametric conjunctive t-rule when combining two BBAs. Assume that the BBAs m_1 and m_2 have been provided by two sensors. (Note that m_1 and m_2 are not assumed to be independent, because we know nothing about the dependence of the two sensors.) Expert knowledge regarding the true value of the variable of interest is represented by BBA m_e, as shown in Table 1. The BBA m_e was artificially constructed by combining m_1 and m_2 using the parametric conjunctive t-rule in which Frank t-norm with parameter $\lambda = 0.5$ is used and both $g_1(x) = x^{0.8}$ and $g_1(x) = e^{x(x-1)}$ are considered.

Table 1 Artificial BBAs defined on $\Omega = \{a, b, c\}$ with different selectable functions

A	$m_1(A)$	$m_2(A)$	$m_e(A)$ with different selectable function $g_1(x)$	
			$g_1(x) = x^{0.8}$	$g_1(x) = e^{x(x-1)}$
ϕ	0	0	0	0
$\{a\}$	0	0	0	0
$\{b\}$	0.3	0.4	0.4858	0.7708
$\{a, b\}$	0	0	0	0
$\{c\}$	0	0	0	0
$\{a, c\}$	0	0	0	0
$\{b, c\}$	0.4	0.4	0.3950	0.2176
Ω	0.3	0.2	0.1192	0.0116

We wish to find an optimal combination rule with Frank t-norm $T_F(\lambda)$ and selectable function $g_1(x) = x^{\beta}$ such that the combination of m_1 and m_2 approaches as close as possible to m_e. Parameter λ varies from 0 to 1 with step size 0.1. The distance between m_e and combination m_{12} is measured by Jousselme's evidential distance [12] defined by

$$d\left(m_{12}, m_e\right) = \sqrt{\frac{1}{2}\left(m_{12} - m_e\right)^T D\left(m_{12} - m_e\right)}, \tag{21}$$

where superscript T denotes matrix transposition and matrix D is defined by

$$D(A, B) = \begin{cases} 1, & A = B = \phi, \\ \dfrac{|A \cap B|}{|A \cup B|}, & \text{otherwise.} \end{cases}$$
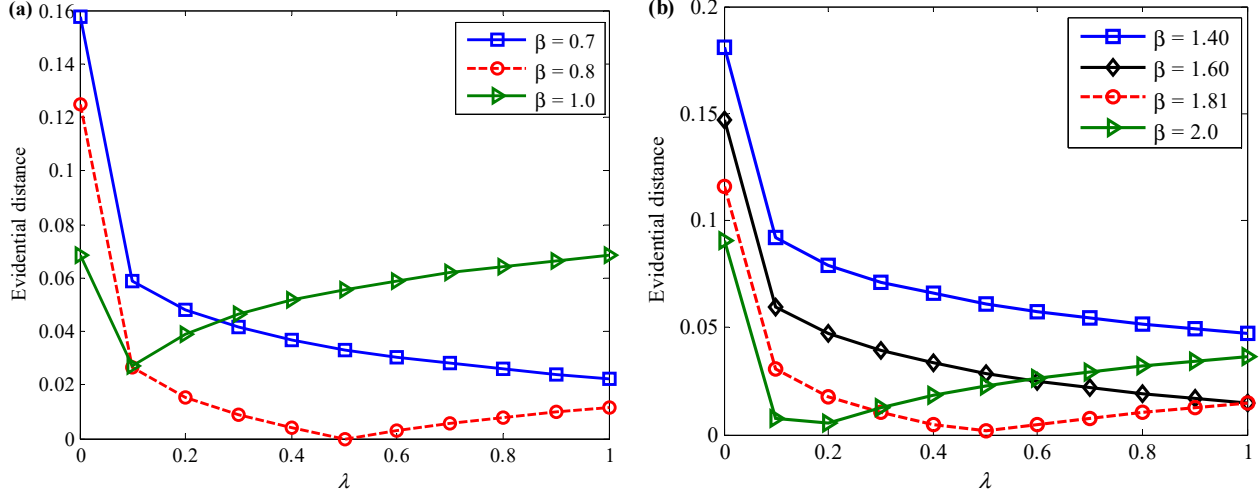
Fig. 3 Evidential distance between m_e and combined BBA m_{12} using PCTR. (a): true combination rule with selectable function $g_1(x) = x^\beta$; (b): true combination rule with selectable function $g_1(x) = e^{x(x-1)}$.

Fig. 3 shows the evidential distance between the artificial BBA and the combination of m_1 and m_2 using PCTR with power function as selectable function. Fig. 3(a) shows that the optimal rule can be exactly found if the selectable function $g_1(x) = x^\beta$ belongs to the same parametric family as the function used to construct the artificial BBA. Alternatively, it can be seen from Fig. 3(b) that the combination using PCTR with appropriate parameter (i.e., $\lambda = 0.5$, β = 1.81) can get very close to the artificial BBA even when expert knowledge is unknown a priori (i.e., a completely unknown selectable function $e^{x(x-1)}$ is used when constructing the artificial BBA). These results show that an optimal rule can be found or approximated through designing selectable functions $g_i(x)$ with appropriate parameters, even when nothing about the true combination rule is known in advance. Note that similar results can be obtained for non-separable BBAs.

4.2 Performance Evaluation

The objective of this section is to show that the performance of the evidential K-NN classifier can be enhanced by using the parametric conjunctive t-rules. In this section, the function $g_1(x) = x^\beta$ and parametric t-norm $\top = T_{DP}(\lambda)$ are considered. Note that function $g_2(x)$ and t-norm $\top'$ cannot play any role in this classification setting, as already remarked in Section 3.2.

To implement the evidential K-NN classifier, practical issues need to be addressed. In particular, parameters k, α, γ_q (q = 1, 2, ..., c), β and $T_{DP}(\lambda)$ need to be preset in advance. There is some literature [34-36, 39] focused on optimizing the parameters k, α and γ_q (q = 1, 2, ..., c). However, the results that will be presented in this section are based on the following simple heuristic method [5]: $\alpha = 0.95$, and γ_q equal to the inverse of the mean Euclidean distance between training patterns belonging to class ω_q. The parameter k was set to 7. Hence, we focus on optimizing parameters λ and β. For convenience, assume that parameter λ takes values in {0, 0.1, 0.2, ..., 1.0}, whereas parameter β varies from 0.5 to 2 by step size 0.1. The optimal parameters λ and β are to maximize the classification accuracy $P(\lambda, \beta)$ defined by

$$P(\lambda, \beta) = \frac{1}{n} \sum_{i=1}^{n} I\left(\omega(\boldsymbol{x}_i) - u_i\right), \tag{22}$$

where u_i is the true class label of object x_i, and $I(.)$ is an indicator function defined by $I(0) = 1$, and $I(x) = 0$ if $x \neq 0$.

Table 2 Data description

No.	Datasets	Abbreviation	Samples	Continuous attributes	Classes
1	Balance-scale	Balance	626	4	3
2	Breast cancer Wisconsin	Breastw	699	9	2
3	Diabetes	Diabetes	768	8	2
4	Ionosphere	Iono	355	34	2
5	Iris	Iris	153	4	3
6	Sonar	Sonar	208	60	2
7	Vehicle	Vehicle	848	18	4
8	Wine	Wine	178	13	3

In the experiments reported in this section, eight UCI datasets with continuous attributes were selected, as shown in Table 2, given the fact that the distance measure $d_i(., .)$ in the evidential K-NN classifier is sensitive to discrete attributes. The precise class labels in these eight UCI datasets were encoded as categorical BBAs m_i, i.e., $m_i(\{\omega_q\}) = 1$ if pattern x_i belongs to class ω_q, otherwise 0.

4.2.1 Experiment 1: Approximate accuracy

This experiment aims to study the performance of the evidential K-NN when parameters (λ, β) take different values. In the experiment, all eight UCI datasets in Table 2 are considered. For each dataset, the Leave-One-Out (LOO) classification accuracy was computed. More precisely, we left each sample out and used the remaining samples to predict the possible label of this sample with a given pair (λ, β); this procedure was repeated until we had tested all samples. The LOO classification accuracy for each pair (λ, β) can be calculated as the ratio of correctly labeled samples to the total number of samples. The conjunctive, cautious and conjunctive t-rules (i.e., PCTR-λ in Table 3) were also implemented under the same conditions.
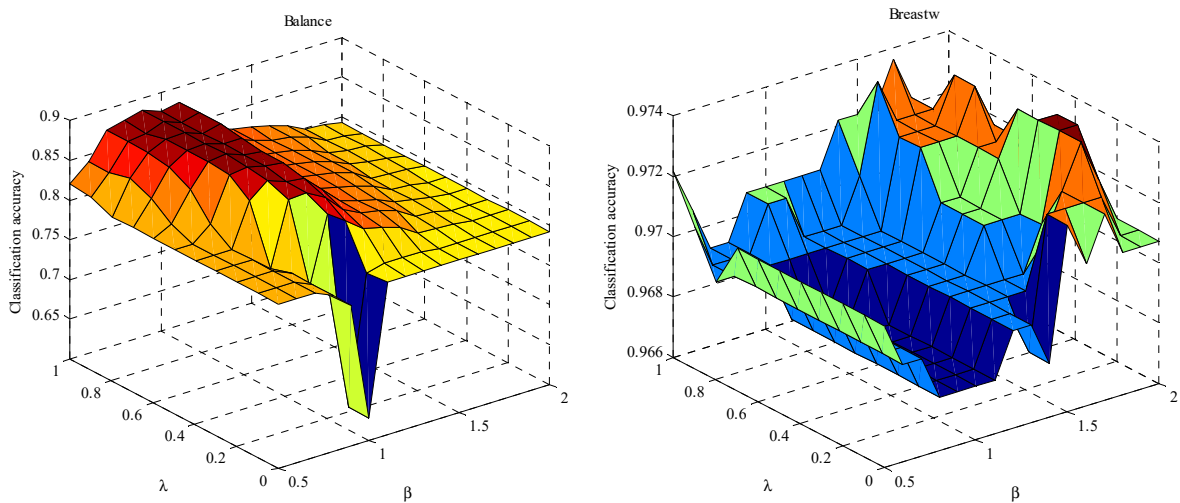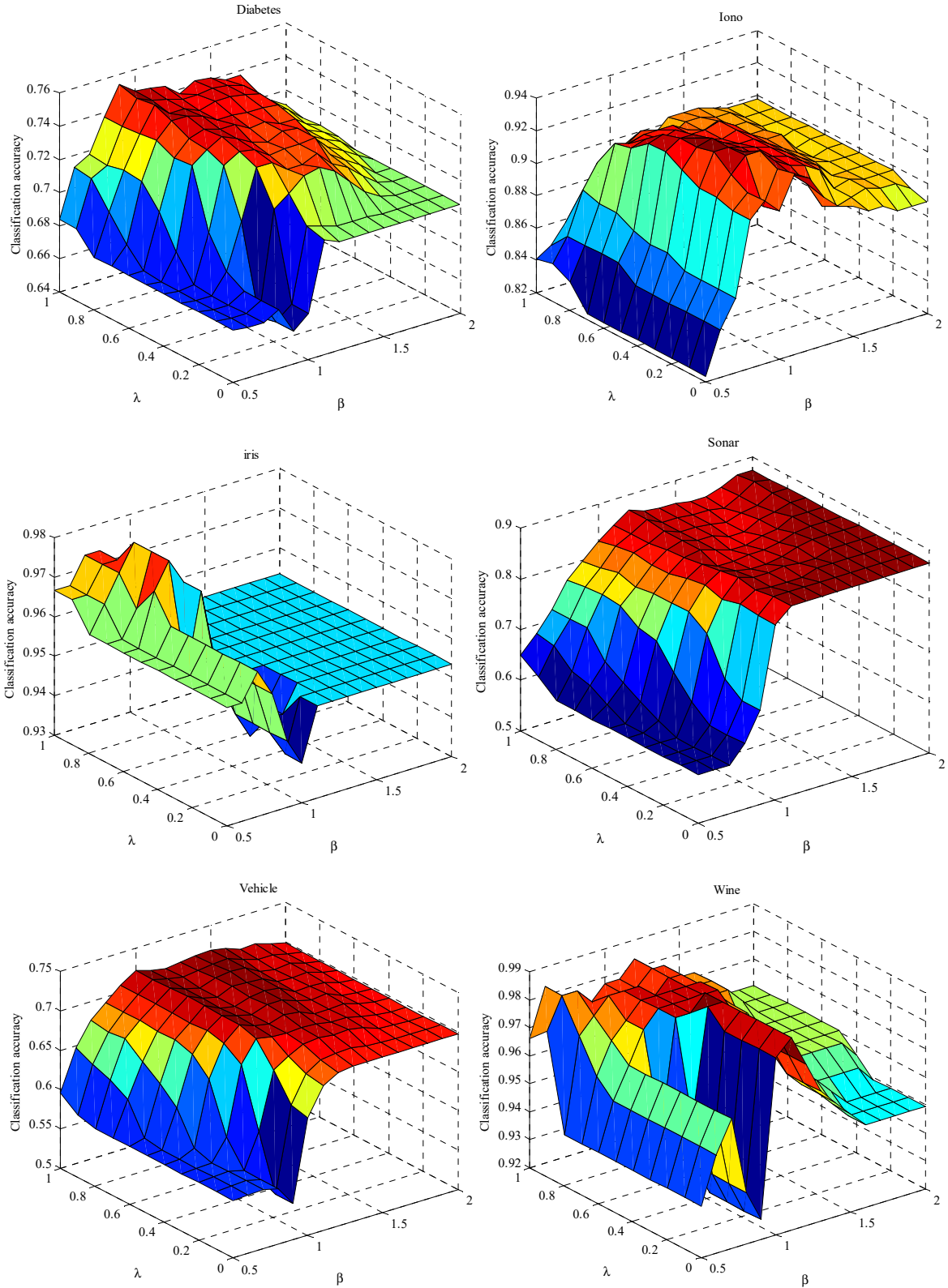
Fig. 4 LOO classification accuracy varying with parameters λ and β

Fig. 4 shows the surfaces of LOO classification accuracy as a function of parameters λ and β. The optimal LOO classification accuracies are presented in Table 3, in which the best results are shown in bold.

From Table 3, it can be seen that the parametric conjunctive t-rules outperform the cautious and conjunctive rules

in most cases. This observation suggests that parametric conjunctive t-rules can improve the performance of the evidential K-NN classifier when optimizing the parameters λ and β simultaneously. From Fig. 4, we can see that there is more than one optimal solution in most cases, and optimal LOO classification accuracy can be achieved when parameter $\beta \neq 1$, which indicates that function $g_1(x)$ has a critical influence on the performance of PCTR.

Table 3 Optimal LOO classification accuracies of the evidential K-NN classifier *1

Datasets	Cautious rule	conjunctive rule	PCTR-λ	PCTR-(λ, β)
Balance	0.6272	0.8767	0.8768	**0.8848**
Breastw	0.9678	0.9707	0.9707	**0.9736**
Diabetes	0.6680	0.7474	0.7487	**0.7539**
Iono	0.9117	0.8917	**0.9345**	**0.9345**
Iris	0.9400	0.9600	0.9600	**0.9800**
Sonar	0.8462	0.8269	0.8510	**0.8798**
Vehicle	0.6489	0.7222	**0.7234**	**0.7234**
Wine	0.9831	0.9775	**0.9888**	**0.9888**

*1. PCTR-λ means that only parameter λ in T_{DP} is optimized when $g_1(x)=x^\beta$ with $\beta = 1$. In this case, PCTR-λ is equivalent to conjunctive t-rule.

4.2.2 Experiment 2: Testing accuracy

This second experiment intends to assess the testing/predictive accuracy of the evidential K-NN classifier. In this experiment, we still considered the eight datasets in Table 2. However, a *double cross-validation loop* strategy [10] was implemented on each dataset in order to avoid bias in the selection procedure. The procedure is summarized in Fig. 5.

Input: Arbitrary dataset S and the ranges of $(\lambda, \beta) = [0, 0.1, …, 1.0]×[0.5, 0.6, …, 2.0] \in R^{11×16}$.

Repetition Loop: For $p = 1$ to n_{REP} % n_{REP} is the number of repetition

    Randomly split S in k_1 non-overlapping folds $S^{(1)}$, …, $S^{(k1)}$ with possible equal size;

    Outer Loop: For $i = 1$ to k_1

    1.   Take $S^{(i)}$ as TEST set, and merge the remaining (k_1-1) folds into $S^{(-i)} = S - S^{(i)}$ as CALIBRATION set;

    2.   Randomly split the calibration set in k_2 non-overlapping folds $C^{(1)}$, …, $C^{(k2)}$ with possible equal size;

    3.   **Inner Loop:** For $j = 1$ to k_2

        1)   Take $C^{(j)}$ as VALIDATION set and the remaining (k_2-1) folds $C^{(-j)} = C - C^{(j)}$ as TRAINING set;

        2)   Calculate $P^{(j)}(\lambda, \beta) \in R^{11×16}$ according to (22) for validation set $C^{(j)}$ based on training set $C^{(-j)}$;

        3)   Select $(\lambda*, \beta*)$ according to the following *standard error-like rule* [37]

$$P(\lambda*, \beta*) > \max\{P(\lambda, \beta)\} - std(P(\lambda, \beta))\Big/\sqrt{|P(\lambda, \beta)|}, \tag{23}$$

        where $P(\lambda, \beta)$ is a column vector by straightening matrix $\bar{P} = k_2^{-1}\sum_{j=1}^{k_2} P^{(j)}(\lambda, \beta)$; |.| is cardinality of P.

    4.   Evaluate the classification accuracy $P^{(i)}(\lambda*, \beta*)$ on the test set $S^{(i)}$;

    5.   Average the k_1 test accuracies $P^{(i)}(\lambda*, \beta*)$;

Output: A group of n_{REP} double cross-validation classification accuracies.

Fig. 5 Double cross-validation algorithm

To implement the double cross-validation algorithm, we set $n_{REP} = 10$, $k_1 = 10$ and $k_2 = 5$, and we obtained a group of cross-validation classification accuracies for each dataset; the mean accuracies with standard deviations *std.* are presented in Table 4. Note that there is sometimes more than one optimal selection according to criterion (23). In this case, the best performance among these optimal selections is considered. Furthermore, the surfaces of classification accuracies varying with λ and β on these datasets are similar as those in Fig. 4; therefore, they are omitted due to space constraints.

From Table 4, it can be seen that the parametric conjunctive t-rules outperform the conjunctive and cautious rules in most cases. For the diabetes and iono datasets, the mean classification accuracy obtained by PCTR-(λ, β) is a little lower than that obtained using PCTR-λ. When we look into the details, we find that the training accuracy using PCTR-(λ, β) is higher than that by using PCTR-λ, which indicates that there may be over-fitting when learning PCTR-(λ, β).

Table 4 Classification accuracies estimated by double cross-validation

Datasets	Cautious rule	Conjunctive rule	PCTR-λ	PCTR-(λ, β)
Balance	0.6313 ± 0.0050	0.8747 ± 0.0048	0.8750 ± 0.0050	**0.8851** ± 0.0050
Breastw	0.9682 ± 0.0007	**0.9712** ± 0.0013	**0.9712** ± 0.0012	**0.9712** ± 0.0022
Diabetes	0.6692 ± 0.0082	0.7377 ± 0.0057	**0.7464** ± 0.0061	0.7441 ± 0.0064
Iono	0.9084 ± 0.0053	0.8904 ± 0.0038	**0.9380** ± 0.0044	0.9352 ± 0.0052
Iris	0.9413 ± 0.0028	0.9520 ± 0.0069	0.9527 ± 0.0080	**0.9700** ± 0.0085
Sonar	0.8403 ± 0.0089	0.8067 ± 0.0142	0.8388 ± 0.0118	**0.8748** ± 0.0107
Vehicle	0.6488 ± 0.0087	0.7123 ± 0.0058	0.7170 ± 0.0058	**0.7316** ± 0.0059
Wine	0.9675 ± 0.0099	0.9771 ± 0.0041	0.9804 ± 0.0071	**0.9821** ± 0.0058

To further investigate the performance of the PCTR-based classifier, we compared it with the popular Evidence-Theoretic k-nearest neighbor classifier with parameter Optimization (ETO) [39], in which the conjunctive rule is used and parameters k and γ_q are optimized rather than fixed. We used the same double cross-validation algorithm as above. As shown in Fig. 6, the PCTR-(λ, β)-base classifier outperforms, or at least is comparable with ETO in most cases even though parameters k and γ_q are fixed.
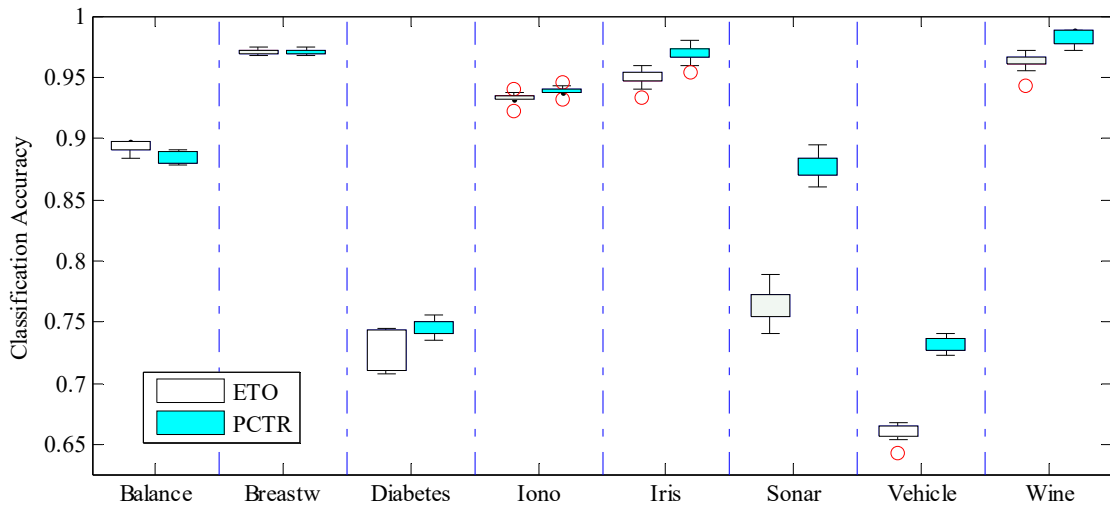

Fig. 6 Comparisons between PCTR-(λ, β) and ETO

Remark 6 When $g_1(x) = x$, the PCTR rule can degenerate to the cautious rule, the conjunctive rule or the conventional conjunctive t-rules. Consequently, the PCTR can be at least as good as these rules. However, we can see that, for instance on the datasets diabetes and iono, the PCTR-λ rule has the best performance. This means that there is some over-fitting, even though a standard error-like method (23) has been adopted. Furthermore, the proposed PCTR may not be able to enhance the performance of the evidential K-NN classifier for all datasets. When the conjunctive (t-)rule is not outperformed by PCTR, it may be possible to find other parametric families of rules that perform better, as mentioned in [22].

5. Conclusions

In this paper, we first introduced a scheme to construct a new family of t-norms with selectable functions satisfying some conditions. Using this family of t-norms, a class of parametric conjunctive t-rules for combining belief functions has been proposed. The parametric conjunctive t-rules, which encompass the cautious and conjunctive rules, can be used to enhance the performance of the evidential K-NN classifier when appropriate selectable functions and tunable parameters are selected.

The performance of the PCTR-based K-NN classifier depends on the choice of the number k of neighbors and on scale parameters γ_q. Hence, optimizing parameters k and γ_q could bring further classification accuracy improvements. Furthermore, we used the power function for $g_1(x)$, but different forms of selectable functions could be investigated as well. Finally, in this paper we optimized parameters (λ, β) using a greedy search procedure, but this strategy is suboptimal. Through global optimization of all parameters in the PCTR rule, it might be possible to further enhance the performance of the evidential K-NN classifier. This research direction is left for future work.

Acknowledgement

Appendix

Proof of Proposition 1

With $x, y \in [1, +\infty]$, we have $g_2(x), g_2(y) \in (0, 1]$. In a similar way, $g_1(x), g_1(y) \in [0, 1]$ can be guaranteed if $x, y \in [0, 1]$. By employing the dual operation $\perp'$ of $\top'$ and negation operation N, we have $x \perp' y = N(N(x) \top' N(y))$. Thus, formula (11) can be rewritten as

$$
x \; \mathcal{T}_{\top \perp'}^{g_i} \; y = \begin{cases} g_1(x) \top g_1(y), & x \vee y \leq 1, \\ g_2^{-1}(g_2(x) \perp' g_2(y)), & x \wedge y > 1, \\ x \wedge y, & otherwise, \end{cases}
\tag{24}
$$

Similar as before, the symbol $\mathcal{T}_{\top \perp'}^{g_i}$ is shorten as $\mathcal{T}$ if no confusions occur in the following proof.

- Commutativity

As can be seen from formula (24), the binary operation $\mathcal{T}$ is defined by using classical norms $\top$, $\perp'$ and $\wedge$. Because of the commutativity of $\top$, $\perp'$ and $\wedge$, operation $\mathcal{T}$ is also commutative.

- **Neutral element**

  If $x \leq 1$, $x \mathcal{T} +\infty = x \wedge +\infty = x$;

  If $x \geq 1$, $x \mathcal{T} +\infty = g_2^{-1}(g_2(x) \perp' g_2(+\infty)) = g_2^{-1}(g_2(x) \perp' 0) = g_2^{-1}(g_2(x)) = x$.

  From the commutativity of $\mathcal{T}$, we have $+\infty \mathcal{T} x = x \mathcal{T} +\infty = x$.

- **Monotonicity**

  Without loss of generality, suppose $w \geq y$, $x \geq z$.

  - Suppose $w \mathcal{T} x = g_1(w) \top g_1(x)$, then $y \mathcal{T} z = g_1(y) \top g_1(z)$ because of $1 \geq w \geq y$, $1 \geq x \geq z$. By the monotonicity of $\top$ and $g_1(.)$, we have $w \mathcal{T} x \geq y \mathcal{T} z$.

  - Suppose $w \mathcal{T} x = w \wedge x$, then there exist $1 \geq w$ or $1 \geq x$, correspondingly, there also exist $1 \geq y$ or $1 \geq z$.

    — If $y \mathcal{T} z = y \wedge z$, we have $w \mathcal{T} x = w \wedge x \geq y \wedge z = y \mathcal{T} z$ because of the monotonicity of $\wedge$;

    — If $y \mathcal{T} z = g_1(y) \top g_1(z)$, we have $g_1(y) \top g_1(z) \leq g_1(y) \wedge g_1(z)$ because the minimum $\wedge$ is the maximum t-norm on $(0, +\infty]$. Without loss of generality, suppose $w \wedge x = x$. We have
      — If $y \leq z$, then $y \wedge z = y$. Because $x \geq z$, we can obtain $x \geq y \wedge z$. Thus, $g_1(w) \wedge g_1(x) \geq g_1(y) \wedge g_1(z) \geq g_1(y) \top g_1(z)$.
      — If $y > z$, then $y \wedge z = z$. and $y \wedge z = z \geq y \top z$. Because $x \geq z$, one has $x \geq y \wedge z$. Thus, $g_1(w) \wedge g_1(x) \geq g_1(y) \top g_1(z)$

  - Suppose $w \mathcal{T} x = g_2^{-1}(g_2(w) \perp' g_2(x))$, then we have $w \wedge x \geq 1$ and $w \mathcal{T} x \geq 1$.

    — If $y \mathcal{T} z = g_2^{-1}(g_2(y) \perp' g_2(z))$, then $w \mathcal{T} x \geq y \mathcal{T} z$ due to the monotonicity of $\perp'$ and function g_2;

    — If $y \mathcal{T} z = y \wedge z$, then $w \mathcal{T} x \geq 1 \geq y \wedge z = y \mathcal{T} z$;

    — If $y \mathcal{T} z = g_1(y) \top g_1(z)$, together with $y \vee z \leq 1$, then $g_1(y) \top g_1(z) \leq 1$. Thus, $w \mathcal{T} x \geq 1 \geq g_1(y) \top g_1(z)$.

  Thus, we have $w \mathcal{T} x \geq y \mathcal{T} z$.

- **Associativity**

  Without loss of generality, suppose $x \geq y \geq z$.

  - If $x \leq 1$, $y \leq 1$, $z \leq 1$, then $(x \mathcal{T} y) \mathcal{T} z = (g_1(x) \top g_1(y)) \top g_1(z) = g_1(x) \top (g_1(y) \top g_1(z)) = x \mathcal{T} (y \mathcal{T} z)$.

  - If $x > 1$, $y > 1$, $z > 1$, we have
    $(x \mathcal{T} y) \mathcal{T} z = g_2^{-1}(g_2(x) \perp' g_2(y)) \mathcal{T} g_2(z) = g_2^{-1}(g_2(g_2^{-1}(g_2(x) \perp' g_2(y))) \perp' g_2(z)) = g_2^{-1}(g_2(x) \perp' g_2(y) \perp' g_2(z))$,
    $x \mathcal{T} (y \mathcal{T} z) = x \mathcal{T} g_2^{-1}(g_2(y) \perp' g_2(z)) = g_2^{-1}(g_2(x) \perp' g_2(g_2^{-1}(g_2(y) \perp' g_2(z)))) = g_2^{-1}(g_2(x) \perp' g_2(y) \perp' g_2(z))$.
    Thus, we have $(x \mathcal{T} y) \mathcal{T} z = x \mathcal{T} (y \mathcal{T} z)$.

  - If $x > 1 > y \geq z$, then we have
    $x \mathcal{T} (y \mathcal{T} z) = x \mathcal{T} (g_1(y) \top g_1(z)) = x \wedge (g_1(y) \top g_1(z)) = g_1(y) \top g_1(z)$, and $(x \mathcal{T} y) \mathcal{T} z = (x \wedge y) \mathcal{T} z = y \mathcal{T} z = g_1(y) \top g_1(z)$. Thus, $(x \mathcal{T} y) \mathcal{T} z = x \mathcal{T} (y \mathcal{T} z)$.

  - If $x \geq y > 1 > z$, then we have
    $x \mathcal{T} (y \mathcal{T} z) = x \mathcal{T} (y \wedge z) = x \mathcal{T} z = x \wedge z = z$, and $(x \mathcal{T} y) \mathcal{T} z = g_2^{-1}(g_2(x) \perp' g_2(y)) \wedge z = z$.
    Thus, $(x \mathcal{T} y) \mathcal{T} z = x \mathcal{T} (y \mathcal{T} z)$。

  - If $x = 1$ or $y=1$ or $z = 1$, suppose $x = 1$, then we have

$x\ \mathcal{T}(y\ \mathcal{T}z) = 1\ \mathcal{T}(y\ \mathcal{T}z) = y\ \mathcal{T}z$, and $(x\ \mathcal{T}y)\ \mathcal{T}z = (1\ \mathcal{T}y)\ \mathcal{T}z = y\ \mathcal{T}z$.

Thus, $(x\ \mathcal{T}y)\ \mathcal{T}z = x\ \mathcal{T}(y\ \mathcal{T}z)$。

Finally, it is evident that $x\ \mathcal{T}y > 0$ because of $x, y \in (0, +\infty]$. ∎

Proof of Proposition 2

The proof is similar as that of Proposition 4.3 in [22]. It is given here for completeness.

Firstly, we validate the premise that the minimum on $(0, +\infty]$ is the largest $\mathcal{T}$. This is evident because any t-norm $\mathcal{T}$ on $(0, +\infty]$ has $+\infty$ as neutral element and is monotonic. Hence, we have $x\ \mathcal{T}y \le x\ \mathcal{T}+\infty = x$ and $x\ \mathcal{T}y \le +\infty\ \mathcal{T}y = y$ for all $x, y \in (0, +\infty]$. So, it results in $x\ \mathcal{T}y \le x \wedge y$. Thus, we have

$$w_{1\ \mathcal{\textcircled{T}}_w 2}(A) \le w_{1\ \textcircled{\wedge}_w 2}(A),\ \ \forall A \subset \Omega.$$

From Lemma 1 in [6], if m is a non-dogmatic BBA with conjunctive weight function w, and if w' is a mapping from $2^\Omega \setminus \{\Omega\} \to (0, +\infty)$ such that $w'(A) \le w(A)$ for all $A \subset \Omega$, then w' is the conjunctive weight function of some non-dogmatic BBA m'.

Consequently, $w_{1\ \textcircled{T}_w 2}$ is a conjunctive weight function since $w_{1\ \textcircled{\wedge}_w 2}$ is a conjunctive weight function. ∎

References

[1] A. Bronevich, I. Rozenberg, The choice of generalized Dempster–Shafer rules for aggregating belief functions, International Journal of Approximate Reasoning 56 (2015) 122–136

[2] M. Cattaneo, Belief functions combination without the assumption of independence of the information sources, International Journal of Approximate Reasoning 52 (3) (2011) 299–315.

[3] M. Chebbah, A. Martin, B.B. Yaghlane, Combining partially independent belief functions, Decision Support Systems, 73 (2015) 37-46.

[4] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Annals of Mathematical Statistics 38 (1967) 325-339.

[5] T. Denoeux, A *k*-nearest neighbor classification rule based on Dempster-Shafer theory, IEEE Transactions on systems, man and cybernetics 25(5) (1995) 804-813.

[6] T. Denoeux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, Artificial intelligence 172(2-3) (2008) 234-264.

[7] T. Denoeux, 40 years of Dempster-Shafer theory, International Journal of Approximate Reasoning 79 (2016) 1-6.

[8] D. Dubois, H. Prade, On the unicity of Dempster's rule of combination, International journal of intelligent systems 1 (1986) 133-142.

[9] Z. Elouedi, K. Mellouli, Pooling dependent expert opinions using the theory of evidence, In: Proceedings of the seventh international conference on information processing and management of uncertainty in knowledge-based systems 1 (1998) 32-39.

[10] P. Filzmoser, B. Liebmann, K. Varmuza, Repeated double cross validation, Journal of Chemometrics 23(4) (2009) 160-171.

[11] L. Jiao, T. Denoeux, Q. Pan, A hybrid belief rule-based classification system based on uncertain training data and expert knowledge, IEEE Transactions on systems, man and cybernetics: systems 46(12) (2016) 1711-1723.

[12] A.-L. Jousselme, D. Grenier, E. Bosse, A new distance between two bodies of evidence, Information fusion 2(2) (2001) 91-101.

[13] A. Kallel, S. Le Hégarat-Mascle, Combination of partially non-distinct beliefs: the cautious-adaptive rule, International Journal of Approximate Reasoning 50(7) (2009) 1000–1021.

[14] E.P. Klement, R. Mesiar, E. Pap, Triangular norms, Position paper 1: basic analytical and algebraic properties, Fuzzy sets and systems 143(1) (2004) 5-26.

[15] C. Lian, S. Ruan, T. Denoeux, Dissimilarity metric learning in the belief function framework, IEEE Transactions on Fuzzy Systems 24(6) (2016) 1555-1564.

[16] Z.G. Liu, Y. Liu, J. Dezert, Q. Pan, Classification of incomplete data based on belief functions and k-nearest neighbors, Knowledge-Based systems 89(2015) 113-125.

[17] Z.G. Liu, Q. Pan, J. Dezert, A new belief-based k-nearest neighbor classification method, Pattern recognition 46 (2013) 834-844.

[18] Z.G. Liu Q. Pan, J. Dezert, G. Mercier, Classification of uncertain and imprecise data based on evidence theory, Neurocomputing 133 (2014) 459-470.

[19] Z.G. Liu, Q. Pan, J. Dezert, G. Mercier, Hybrid Classification System for Uncertain Data, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2016. DOI: 10.1109/TSMC.2016.2622247.

[20] Z.G. Liu, Q. Pan, J. Dezert, A. Martin, Combination of classifiers with optimal weight based on evidential reasoning, IEEE Transactions on Fuzzy Systems, 2017. DOI: 10.1109/TFUZZ.2017.2718483

[21] T. Nakama, E. Ruspini, Combining dependent evidential bodies that share common knowledge, International Journal of Approximate Reasoning 55(9) (2014) 2109–2125.

[22] F. Pichon, Belief functions: Canonical Decompositions and Combination Rules, PhD thesis, Université de Technologie de Compiègne, France, 2009.

[23] F. Pichon, T. Denoeux, T-norm and Uninorm-based combination of belief functions, In: Proceedings of the North American Fuzzy Information Processing Society International Conference, New York, USA, 2008, pp. 19-22.

[34] F. Pichon, T. Denoeux, The unnormalized Dempster's rule of combination: a new justification from the least commitment principle and some extensions, Journal of automated reasoning 45(1) (2010) 61-87.

[25] B. Quost, M.-H. Masson, T. Denœux, Classifier fusion in the Dempster–Shafer framework using optimized t-norm based combination rules, International Journal of Approximate Reasoning 52(3) (2011) 353–374.

[26] G. Shafer, A mathematical theory of evidence, Princeton University Press, Princeton, N.J., 1976.

[27] G. Shafer, A mathematical theory of evidence turn 40, International Journal of Approximate Reasoning 79 (2016) 7-25.

[28] G. Shafer, Dempster's rule of combination, International Journal of Approximate Reasoning 79 (2016) 26-40.

[29] G. Shafer, The problem of dependent evidence, International Journal of Approximate Reasoning 79 (2016) 41-44.

[30] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, International Journal of Approximate Reasoning 9(1) (1993) 1-35.

[31] P. Smets, The canonical decomposition of a weighted belief, In: Proceedings of the 14th international joint conference on artificial intelligence, San Mateo, California, USA, 1995, pp. 1896-1901.

[32] P. Smets, R. Kennes, The transferable belief model, Artificial intelligence 66 (1994) 191-243.

[33] X. Su, M. Sankaran, W. Han, Y. Deng, Combining dependent bodies of evidence, Applied Intelligence 44(3) (2016) 634-644.

[34] Z.G. Su, P.H. Wang, A robust adaptive version of evidence theoretic k-NN classification rule, In: Proceedings of the 6th International conference on Fuzzy Systems and Knowledge discovery 4 (2009) 525-529.

[35] Z.G. Su, P.H. Wang, Improved adaptive evidential k-NN rule and its application for monitoring level of coal powder filling in ball mill, Journal of process control 19(10) (2009) 1751-1762.

[36] Z.G. Su, P.H. Wang, X.J. Yu, Immune genetic algorithm-based adaptive evidential model for estimating unmeasured parameter: Estimating levels of coal powder filling in ball mill, Expert Systems With Applications 37(7) (2010) 5246-5258.

[37] K. Varmuza, P. Filzmoser, M. Hilchenbach, H. Krüger, J. Silén, KNN classification-evaluated by repeated double cross validation: Recognition of minerals relevant for comet dust, Chemometrics and Intelligent Laboratory Systems 138 (2014) 64-71.

[38] Q. Wang, Z.G. Su, B. Rezaee, P.H. Wang, Constructing T-S fuzzy model from imprecise and uncertain knowledge represented as fuzzy belief functions, Neurocomputing 166 (2015) 319-336.

[39] L.M. Zouhal, T. Denoeux, An evidence-theoretic k-NN rule with parameter optimization, IEEE Transactions on systems, man and cybernetics-Part C 28(2) (1998) 263-271.