The Use of Fuzzy t-Conorm Integral for Combining Classifiers*

David Štefka and Martin Holeňa

Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic
{stefka,martin}@cs.cas.cz

Abstract. Choquet or Sugeno fuzzy integrals are commonly used for aggregating the results of different classifiers. However, both these integrals belong to a more general class of fuzzy t-conorm integrals. In this paper, we describe a framework of a fuzzy t-conorm integral and its use for combining classifiers. We show the advantages of this approach to classifier combining in several benchmark tests.

Keywords: combining classifiers, classifier aggregation, fuzzy integral, t-conorm integral, fuzzy measure.

1 Introduction

Combining several different classifiers in order to improve the quality of classification is a common approach today and many different methods have been described in the literature. Out of these, fuzzy integral is often used. Although there is a general class of fuzzy t-conorm integrals, introduced by Murofushi and Sugeno [1], only two particular integrals (Choquet integral and Sugeno integral) from this class are commonly used.

In this paper, we describe a framework of fuzzy t-conorm integral, investigate which particular types of fuzzy t-conorm systems are useful for combining classifiers, and perform tests on real data to show the performance of combining classifiers by using fuzzy t-conorm integral.

The paper is structurred as follows: in Section 2, we introduce the formalism needed for combining classifiers. Section 3 deals with fuzzy integrals and fuzzy measures. Subsection 3.1 introduces the fuzzy t-conorm integral, Subsection 3.2 describes the way a fuzzy t-conorm integral can be used for combining classifiers, and Subsection 3.3 investigates different t-conorm systems for integration and considers which particular systems are useful for classifier combining. Section 4 contains experimental results, and finally, Section 5 then concludes the paper.

* The research reported in this paper was partially supported by the Program "Information Society" under project 1ET100300517 (D. Štefka) and the grant No. 201/05/0325 of the Grant Agency of the Czech Republic (M. Holeňa), and partially supported by the Institutional Research Plan AV0Z10300504.

K. Mellouli (Ed.): ECSQARU 2007, LNAI 4724, pp. 755–766, 2007.

2 Combining Classifiers

We define a classifier as a mapping $\phi : \mathcal{X} \rightarrow [0,1]^N$, where $\mathcal{X} \subseteq \mathbb{R}^n$ is a n-dimensional feature space, and $\phi(\boldsymbol{x}) = [\mu_1(\boldsymbol{x}), \ldots, \mu_N(\boldsymbol{x})]$ is a vector of "weights of classification" of the pattern $\boldsymbol{x}$ to each class $C_1, \ldots, C_N$. This type of classifier is called *measurement classifier* [2] or *possibilistic classifier* [3] in the literature.

Whenever we want to use classifier combining to improve the quality of the classification, first of all, we need to create a team of diverse classifiers $\phi_1, \ldots, \phi_k$. Most often, the team consists of classifiers of the same type, which differ only in their parameters, dimensionality, or training sets – such a team is usually called an *ensemble* of classifiers. The restriction to classifiers of the same type is not essential, but it ensures that the classifiers' outputs are consistent. Well-known methods for ensemble creation are *bagging* [4], *boosting* [5], or *multiple feature subset* (MFS) methods [6].

After we have constructed an ensemble of classifiers $\phi_1, \ldots, \phi_k$, we have to use some function $\mathcal{A}$ to aggregate the results of the individual classifiers to get the final prediction, i.e. $\Phi(\mathbf{x}) = \mathcal{A}(\phi_1(\mathbf{x}), \ldots, \phi_k(\boldsymbol{x}))$, where Φ is the final aggregated classifier. The output of an ensemble can be structured to a $k \times N$ matrix, called *decision profile* (DP):

$$DP(\boldsymbol{x}) = \begin{pmatrix} \phi_1(\boldsymbol{x}) \\ \phi_2(\boldsymbol{x}) \\ \vdots \\ \phi_k(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} \mu_{1,1}(\boldsymbol{x}) & \mu_{1,2}(\boldsymbol{x}) & \ldots & \mu_{1,N}(\boldsymbol{x}) \\ \mu_{2,1}(\boldsymbol{x}) & \mu_{2,2}(\boldsymbol{x}) & \ldots & \mu_{2,N}(\boldsymbol{x}) \\ & & \ddots & \\ \mu_{k,1}(\boldsymbol{x}) & \mu_{k,2}(\boldsymbol{x}) & \ldots & \mu_{k,N}(\boldsymbol{x}) \end{pmatrix}, \tag{1}$$

The $i-$th row of the $DP(\boldsymbol{x})$ is an output of the corresponding classifier ϕ_i, and the $j-$th column contains the weights of classification of $\boldsymbol{x}$ to the corresponding class C_j given by all the classifiers.

Many methods for aggregating the ensemble of classifiers into one final classifier have been reported in the literature. A good overview of the commonly used aggregation methods can be found in [3]. These methods comprise simple arithmetic rules (voting, sum, product, maximum, minimum, average, weighted average, cf. [3,7]), fuzzy integral [3,8], Dempster-Shafer fusion [3,9], second-level classifiers [3], decision templates [3], and many others.

In this paper, we deal with the fuzzy integral for classifier combining. In the literature, the Sugeno or Choquet integral are commonly used for classifier aggregation, but there is also a more general framework for fuzzy integral by Murofushi and Sugeno [1], which contains Sugeno and Choquet integral as a special case. This framework is called *fuzzy t-conorm integral*, and is described in the next section.

3 Fuzzy Integral and Fuzzy Measure

Through the rest of the paper, we use the following notation for common t-norms and t-conorms:

– Standard: $x \wedge_S y = \min(x, y), x \vee_S y = \max(x, y)$
– Łukasiewicz: $x \wedge_L y = \max(x + y - 1, 0), x \vee_L y = \min(x + y, 1)$
– Product: $x \wedge_P y = xy, x \vee_P y = x + y - xy$
– Drastic: $x \wedge_D y = x$ if $y = 1$, y if $x = 1$, and 0 otherwise; $x \vee_D y = x$ if $y = 0$, y if $x = 0$, and 1 otherwise.

In general, fuzzy integrals (see [10] for details) can be looked upon as aggregation operators with respect to a fuzzy measure. The integrand (function values to aggregate) is integrated with the the fuzzy measure (expressing the importance of individual elements). A fuzzy measure is a generalization of the classical probability measure (the difference is that it does not need to fulfil the conditions of σ-additivity), and can be defined as follows:

Definition 1. *Let X be a nonempty set, Ω a set of subsets of X, such that $\emptyset, X \in \Omega$. A* fuzzy measure *over (X, Ω) is a function $g : \Omega \to [0, 1]$, such that:*

1. *$g(\emptyset) = 0, g(X) = 1$, and*
2. *if $A, B \in \Omega, A \subseteq B$, then $g(A) \leq g(B)$.*

The tuple (X, Ω, g) is called a fuzzy measure space.

If $X = \{x_1, \ldots, x_k\}$ is a finite set with k elements and Ω is the power set (set of all subsets) of X, then the fuzzy measure is determined by its 2^k values. Sugeno introduced the so-called *λ-fuzzy measure* [11], which needs only k values $g(x_1) = g_1, \ldots, g(x_k) = g_k$ to be determined properly (these values are called *fuzzy densities*), and the remaining values are computed using

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B), \tag{2}$$

where $A, B \in \Omega, A \cap B = \emptyset$, and λ is the only non-zero (if the fuzzy densities do not sum to one; if they do, $\lambda = 0$) root greater than -1 of the equation

$$\lambda + 1 = \prod_{i=1}^{k} (1 + \lambda g_i). \tag{3}$$

After we have defined a fuzzy measure, we can define the two commonly used fuzzy integrals. These are Choquet fuzzy integral and Sugeno fuzzy integral. Since for classifier combining X is a finite set, we restrict ourselves to so-called *simple functions*:

Definition 2. *A function $f : X \to [0, 1]$ is* simple*, if there exist $n \in \mathbb{N}$, $a_1, \ldots, a_n \in [0, 1], a_1 \leq a_2 \leq \cdots \leq a_n$, and $D_1, \ldots, D_n \subseteq X, D_i \cap D_j = \emptyset$ for $i \neq j$, such that $\forall x \in X$:*

$$f(x) = \sum_{i=1}^{n} a_i \mathbf{1}_{D_i}(x), \tag{4}$$

where $\mathbf{1}_{D_i}$ is the characteristic function of D_i.

Equivalent form of (4) is:

$$f(x) = \sum_{i=1}^{n} (a_i - a_{i-1}) \mathbf{1}_{A_i}(x), \tag{5}$$

where $a_0 = 0$ *and* $A_i = \bigcup_{j=i}^{n} D_j$.

Definition 3. *Let g be a fuzzy measure from Def. 1, f a simple function with a_i and $A_i \in \Omega, i = 1, \dots n$ from Def. 2. Then the* Sugeno integral *of f with respect to a fuzzy measure g, denoted as $\int_S f \mathrm{d}g$, is defined by:*

$$\int_S f \mathrm{d}g = \bigvee_{S\ i=1}^{n} (a_i \wedge_S g(A_i)). \tag{6}$$

Definition 4. *Let g be a fuzzy measure from Def. 1, f be a simple function with a_i and $A_i \in \Omega, i = 1, \dots n$ from Def. 2. Then the* Choquet integral *of f with respect to a fuzzy measure g, denoted as $\int_C f \mathrm{d}g$, is defined by:*

$$\int_C f \mathrm{d}g = \sum_{i=1}^{n} (a_i - a_{i-1}) g(A_i). \tag{7}$$

3.1 The Fuzzy t-Conorm Integral

Although Sugeno and Choquet fuzzy integrals are used routinely in many applications, they belong to the class of the so-called *fuzzy t-conorm integrals*, which were introduced in [1]. In this section, we present the framework of t-conorm fuzzy integral for simple functions, following [10]. For details and further information about the following definitions, refer to [1,10].

The individual types of fuzzy t-conorm integrals differ in the way how they bind together the spaces of integrand, measure, and integral. To formalize this, the spaces are linked together by the following definition.

Definition 5. *Let $\triangle, \perp, \underline{\perp}$ be continuous t-conorms, each of which is either Archimedean, or $\vee_S$. Let $([0,1], \triangle), ([0,1], \perp)$, and $([0,1], \underline{\perp})$ denote the spaces of values of integrand, measure, and integral, respectively. Let $\odot : ([0,1], \triangle) \times ([0,1], \perp) \to ([0,1], \underline{\perp})$ be a non-decreasing operation in both variables satisfying the following:*

1. *$\odot$ is continuous on $(0,1]^2$,*
2. *$a \odot x = 0$ if and only if $a = 0$ or $x = 0$,*
3. *If $x \perp y < 1$, then $\forall a \in [0,1] : a \odot (x \perp y) = (a \odot x) \underline{\perp} (a \odot y)$,*
4. *If $a \triangle b < 1$, then $\forall x \in [0,1] : (a \triangle b) \odot x = (a \odot x) \underline{\perp} (b \odot x)$.*

Then $\mathcal{F} = (\triangle, \perp, \underline{\perp}, \odot)$ is called a t-conorm system for integration. *If all the three t-conorms $\triangle, \perp, \underline{\perp}$ are Archimedean, $\mathcal{F}$ is then called* Archimedean.

Prior to defining the t-conorm integral for simple functions, we have to define the *pseudo-difference*.

Definition 6. *Let $\triangle$ be a t-conorm. An operation $-_\triangle : [0,1]^2 \to [0,1]$, defined as*

$$a -_\triangle b = \inf\{c | b \triangle c \geq a\}, \tag{8}$$

is called pseudo-difference *of a and b with respect to $\triangle$.*

Remark 1. Pseudo-difference is dual to the residue $\Rightarrow_\wedge$, defined as $a \Rightarrow_\wedge b = \sup\{c | a \wedge c \leq b\}$, $\wedge$ being a t-norm.

Example 1. If $\triangle = \vee_S$, then

$$a -_{\vee_S} b = \begin{cases} a & \text{if } a \geq b \\ 0 & \text{if } a < b \end{cases} \tag{9}$$

Example 2. If $\triangle = \vee_L$, then

$$a -_{\vee_L} b = \max(a - b, 0) \tag{10}$$

Now we are ready to define fuzzy t-conorm integral.

Definition 7. *Let (X, Ω, g) be a fuzzy measure space, $\mathcal{F} = (\triangle, \perp, \underline{\perp}, \odot)$ a t-conorm system for integration and f a simple function with a_i and $A_i, i = 1, \dots n$ from Def. 2. The* fuzzy t-conorm integral *of f based on $\mathcal{F}$ with respect to g is defined by:*

$$\int_{\mathcal{F}} f \odot \mathrm{d}g = \underline{\perp}_{i=1}^{n}((a_i -_\triangle a_{i-1}) \odot g(A_i)). \tag{11}$$

Example 3. For $\mathcal{F} = (\vee_S, \vee_S, \vee_S, \wedge_S)$, we get the Sugeno integral.

Example 4. For $\mathcal{F} = (\vee_L, \vee_L, \vee_L, \cdot)$, where $\cdot$ is the ordinary multiplication, we get the Choquet integral.

3.2 Using Fuzzy Integral for Classifier Aggregation

Suppose we have a team of classifiers $\phi_1, \dots, \phi_k : \mathcal{X} \to [0,1]^N$ and a t-conorm system for integration $\mathcal{F} = (\triangle, \perp, \underline{\perp}, \odot)$. For a given pattern $\boldsymbol{x} \in \mathcal{X}$, we organize the outputs of the classifiers in the team to a decision profile (1) $DP(\boldsymbol{x})$, and for each class $C_j, j = 1, \dots, N$, we fuzzy-integrate the j-th column of $DP(\boldsymbol{x})$, resulting in the aggregated weight of the classification of $\boldsymbol{x}$ to the class C_j.

In this case, the space (universe) X from Def. 1 and 2 is the set of all classifiers, i.e. $X = \{\phi_1, \dots, \phi_k\}$, and the j-th column of $DP(\boldsymbol{x})$ is a simple function $f : X \to ([0,1], \triangle)$.

To obtain $a_i, D_i, A_i, i = 1, \dots, k$ from Def. 2, we sort the values in j-th column of $DP(\boldsymbol{x})$ in ascending order, and we denote the sorted values $a_1, \dots, a_k$. In other words, a_1 is the lowest weight of classification of $\boldsymbol{x}$ to C_j (acquired by some classifier, which we will denote $\phi_{(1)}$), a_2 the second lowest (acquired by $\phi_{(2)}$), and so on.

Using this notation, $f(\phi_{(i)}) = a_i$, and moreover

$$f(x) = \sum_{i=1}^{k} a_i \mathbf{1}_{D_i}(x) = \sum_{i=1}^{k} (a_i - a_{i-1}) \mathbf{1}_{A_i}(x), \tag{12}$$

where $x \in X = \{\phi_1, \dots, \phi_k\}$, and $a_0 = 0 \leq a_1 \leq a_2 \leq \dots \leq a_k$, $D_i = \{\phi_{(i)}\}$, $A_i = \bigcup_{l=i}^{k} D_l$, which corresponds to Def. 2.

So far, we have properly determined the function f to integrate (integrand), and the space of integrand values $([0,1], \triangle)$. Now we need to define a fuzzy measure g. Again, let $X = \{\phi_1, \dots, \phi_k\}$ be the universe, and let Ω be the power set of X, and $g : \Omega \to ([0,1], \perp)$ a mapping satisfying Def. 1. In the rest of the paper, we use the λ-fuzzy measure, but any other fuzzy measure could be used.

As can be seen from (11), we do not need all the 2^k values of g for the integration – only $g(A_i), i = 1, \dots, k$ are needed. Let g be a λ-fuzzy measure, defined in Section 3 (replacing g_i by the permuted $g_{(i)}$), where $g_{(i)} = g(\phi_{(i)}) \in [0,1], i = 1, \dots, k$ represent the importance of individual classifiers. These values are called *fuzzy densities*, and can be defined for example as $g_{(i)} = 1 - Err(\phi_i)$, where $Err(\phi_i)$ denotes the error rate of classifier ϕ_i. To compute all the necessary values of g, we use the following recursive approach based on (2), for $l = k, \dots, 3$:

$$\begin{aligned} g(A_k) &= g(D_k) = g(\{\phi_{(k)}\}) = g_{(k)} \\ g(A_{l-1}) &= g(D_{l-1} \cup A_l) = g(\{\phi_{(l-1)}, \dots, \phi_{(k)}\}) = \\ &= g(A_l) + g_{(l-1)} + \lambda g(A_l) g_{(l-1)} \\ &\dots \\ g(A_1) &= g(D_1 \cup A_2) = g(\{\phi_{(1)}, \dots, \phi_{(k)}\}) = \\ &= g(A_2) + g_{(1)} + \lambda g(A_2) g_{(1)} \end{aligned} \tag{13}$$

Now, having properly identified the function to integrate f and the fuzzy measure g, we can finally compute the fuzzy t-conorm integral of f based on $\mathcal{F}$, with respect to g, according to (11). The result of the integration is the aggregated weight of the classification of $\boldsymbol{x}$ to the class C_j. The whole process is summarized in Fig. 1.

3.3 Classification of t-Conorm Systems

The framework of fuzzy t-conorm integral provides many different types of fuzzy integrals, depending on the t-conorm system for integration used. However, not all combinations of t-conorms give t-conorm systems for integration, and moreover, not all t-conorm systems for integration provide useful approach to combining classifiers. In this section, we classify the t-conorm systems into classes, and investigate each class in detail.

One important class of t-conorm systems are Archimedean systems, i.e. $\mathcal{F} = (\triangle, \perp, \underline{\perp}, \odot)$, where all the t-conorms are Archimedean. As noticed in [10], if $\mathcal{F}$ is Archimedean, the corresponding fuzzy integral can be expressed as Choquet integral (and hence it has nearly the same properties as the Choquet integral).

Input: Team of classifiers $\phi_1, \ldots, \phi_k$, t-conorm system for integration $\mathcal{F}$, fuzzy densities $g_1, \ldots, g_k$ (e.g. $g_i = 1 - Err(\phi_i)$), pattern $\boldsymbol{x}$ to classify.
Output: $\Phi(\boldsymbol{x}) = [\mu_1(\boldsymbol{x}), \ldots, \mu_N(\boldsymbol{x})]$, i.e. a vector of weights of classification of $\boldsymbol{x}$ to all the classes $C_1, \ldots, C_N$.

1. Let each of the classifiers $\phi_1, \ldots, \phi_k$ predict independently, resulting in a decision profile $DP(\boldsymbol{x})$, of the form (1).
2. If $\sum_{i=1}^{k} g_i = 1$, set $\lambda = 0$, otherwise calculate λ as the only non-zero root greater than -1 of the polynomial equation (2).
3. For $j = 1, \ldots, N$ aggregate the j-th column of $DP(\boldsymbol{x})$ (i.e. weights of classification of $\boldsymbol{x}$ to C_j from all the classifiers) using fuzzy t-conorm integral:
   (a) Sort the values in the j-th column of $DP(\boldsymbol{x})$ in ascending order, denoting the sorted values $a_1, \ldots, a_k$. The corresponding classifiers will be denoted $\phi_{(1)}, \ldots, \phi_{(k)}$, and the corresponding fuzzy densities $g_{(1)}, \ldots, g_{(k)}$.
   (b) Using (13), compute $g(A_i), i = 1, \ldots, k$, where $A_i = \{\phi_{(i)}, \ldots, \phi_{(k)}\}$.
   (c) Using (11), compute the aggregated weight of classification of $\boldsymbol{x}$ to class C_j:

$$\mu_j(\boldsymbol{x}) = \int_{\mathcal{F}} f \odot \mathrm{d}g = \underline{\perp}_{i=1}^{n}((a_i -_{\triangle} a_{i-1}) \odot g(A_i)). \tag{14}$$

4. End with output $\Phi(\boldsymbol{x}) = [\mu_1(\boldsymbol{x}), \ldots, \mu_N(\boldsymbol{x})]$.

Fig. 1. Aggregation of classifier team using fuzzy t-conorm integral with respect to a λ-fuzzy measure

Let $h_{\triangle}, h_{\perp}, h_{\underline{\perp}}$ denote the generators of $\triangle, \perp, \underline{\perp}$, respectively, then the following holds:

$$\int_{\mathcal{F}} f \odot \mathrm{d}g = h_{\underline{\perp}}^{-1}\left[\min\left(h_{\underline{\perp}}(1), \int_C h_{\triangle} \circ f \mathrm{d}(h_{\perp} \circ g)\right)\right], \tag{15}$$

where $\circ$ denotes function composition. Therefore, we will focus our attention to non-Archimedean t-conorm systems in the rest of the paper.

From the non-Archimedean t-conorm systems (i.e. at least one of the t-conorms is $\vee_S$, and the rest is Archimenean), we will set aside systems with $\triangle \neq \underline{\perp}$. The reason for this is that if integral is regarded as mean value of integrands, then the spaces of integrand $([0,1], \triangle)$ and integral $([0,1], \underline{\perp})$ must be the same, i.e. $\triangle = \underline{\perp}$ (see [1,10] for details). However, this class of t-conorm systems is not as large as it may seem, because of the following proposition.

Proposition 1. *Let $\mathcal{F} = (\triangle, \perp, \underline{\perp}, \odot)$ be a non-Archimedean t-conorm system for integration, such that $\triangle = \underline{\perp}$, and $\odot$ is not constant in the second argument on (0,1]. Then $\triangle = \perp = \underline{\perp} = \vee_S$.*

Proof. Since $\mathcal{F}$ is not Archimedean, i.e. at least one of $\triangle, \perp, \underline{\perp}$ is $\vee_S$ and the rest is Archimedean, and $\triangle = \underline{\perp}$, there are only two situations possible:

– $\mathcal{F} = (\triangle, \vee_S, \triangle, \odot)$, where $\triangle$ is $\vee_S$, or continuous Archimedean t-conorm. Suppose that $\triangle \neq \vee_S$, i.e. it is continuous and Archimedean. According to Req. 3 from Def. 5, when $x \vee_S y < 1$, then $\forall a \in [0,1]$

$$a \odot (x \vee_S y) = (a \odot x) \triangle (a \odot y). \tag{16}$$

Let $x = y < 1$. Then (16) reduces to

$$a \odot x = (a \odot x) \triangle (a \odot x). \tag{17}$$

This conflicts with the fact that $\triangle$ is Archimedean, i.e. $u \triangle u > u\ \forall u \in (0,1)$.

- $\mathcal{F} = (\vee_S, \perp, \vee_S, \odot)$, where $\perp$ is $\vee_S$, or continuous Archimedean t-conorm. Suppose that $\perp \neq \vee_S$, i.e. it is continuous and Archimedean. According to Req. 3 from Def. 5, when $x \perp y < 1$, then $\forall a \in [0,1]$

$$a \odot (x \perp y) = (a \odot x) \vee_S (a \odot y). \tag{18}$$

Let $x = y$, such that $x \perp x < 1$. Then (18) reduces to

$$a \odot (x \perp x) = a \odot x. \tag{19}$$

Since $\perp$ is Archimedean, this is with conflict with the fact that $\odot$ is not constant in the second argument on (0,1].

In both cases $\triangle = \perp = \underline{\perp} = \vee_S$, which proves the proposition. □

We could still construct t-conorm systems for integration of the form $\mathcal{F} = (\vee_S, \perp, \vee_S, \odot)$, with $\odot$ constant in the second argument (because then (19) holds), i.e., in fact, integrals with no respect to the measure. However, this is not very useful. Fuzzy t-conorm systems of the form $\mathcal{F} = (\vee_S, \vee_S, \vee_S, \odot)$ are called $\vee_S$*-type* systems. The following proposition expresses that for $\vee_S$-type systems, Req. 3 and 4 from Def. 5 are satisfied automatically.

Proposition 2. *Let $\odot : [0,1] \times [0,1] \to [0,1]$ be a non-decreasing operation satisfying requirements 1 and 2 from Def. 5. Then $\mathcal{F} = (\vee_S, \vee_S, \vee_S, \odot)$ is a t-conorm system for integration.*

Proof. We have to prove the Req. 3 and 4 from Def. 5. Since the proof of the latter is analogous to the proof of the former, we will prove only Req. 3, i.e.: when $x \vee_S y < 1$, then $\forall a \in [0,1] : a \odot (x \vee_S y) = (a \odot x) \vee_S (a \odot y)$. Without loss of generality, we can assume that $x \leq y$. This implies $x \vee_S y = y < 1$, and $a \odot (x \vee_S y) = a \odot y$. Since $\odot$ is non-decreasing, $(a \odot x) \leq (a \odot y)$, thus $(a \odot x) \vee_S (a \odot y) = a \odot y$, which proves the proposition. □

Among $\vee_S$-type systems, *quasi-Sugeno* systems, for which $\odot = \wedge$, $\wedge$ being a t-norm, play an important role. However, not all t-norms can be used:

Proposition 3. *$\mathcal{F} = (\vee_S, \vee_S, \vee_S, \wedge)$, where $\wedge$ is a t-norm, is a t-conorm system for integration if and only if $\wedge$ is continuous on $(0,1]^2$ and without zero divisors.*

Proof. Recall that an element $a \in (0,1)$ is called a zero divisor of a t-norm $\wedge$ if there exists some $x \in (0,1)$, such that $a \wedge x = 0$. The implication $\Rightarrow$ is trivial. The other implication can be proved using Prop. 2 (with taking in mind that $a \wedge 0 = 0$ for any t-norm $\wedge$). □

For example the Łukasiewicz or the drastic t-norms have zero divisors and hence cannot be used in quasi-Sugeno systems. We can summarize the previous to create the following classification of t-conorm systems for integration:

1. Archimedean systems – can be expressed using Choquet integral, with Choquet integral as a special case.
2. Non-Archimedean systems
   (a) Systems with $\triangle = \underline{\perp}$ – lead to $\vee_S$-type systems, with quasi-Sugeno systems as a special case (and Sugeno integral in particular).
   (b) Systems with $\triangle \neq \underline{\perp}$ – do not express mean value of integrand.

4 Experiments

In this section, we present results of our experiments with quasi-Sugeno t-conorm systems. The first experiment shows the advantage of quasi-Sugeno integral over Sugeno integral, because we can fine-tune the t-norm $\wedge$ for particular data. The second experiment studies performance of quasi-Sugeno integral for "easy" data.

4.1 Experiment 1

In this scenario, we used quasi-Sugeno integral with Aczel-Alsina t-norm, i.e. a t-conorm system $\mathcal{F} = (\vee_S, \vee_S, \vee_S, \wedge_\alpha^{AA})$, and compared its performance for different parameters of the t-norm on two datasets. The Aczel-Alsina t-norm [12] with parameter α is defined as follows:

$$x \wedge_\alpha^{AA} y = \begin{cases} x \wedge_D y & \text{if } \alpha = 0 \\ x \wedge_S y & \text{if } \alpha = \infty \\ \exp(-((-\log x)^\alpha + (-\log y)^\alpha)^{1/\alpha}) & \text{if } \alpha \in (0, \infty) \end{cases} \tag{20}$$

For our experiment, we chose $\alpha = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.2, 1.5, 2, 3$, and $\alpha = \infty$ (which corresponds to the Sugeno integral). Since $\wedge_D$ is not continuous on $(0, 1]^2$, $\alpha = 0$ can not be used. The classifiers $\phi_1, \ldots, \phi_k$ were Bayesian classifiers [13], which used all possible subsets of features (i.e. all possible 1-D Bayesian classifiers, 2-D, 3-D, and so on). The ensemble was aggregated according to the algorithm in Fig. 1.

The results of the testing on two bechmark datasets, Phoneme (from the Elena database, [14]) and Pima (from the UCI repository, [15]) datasets, are shown in Fig. 2 and 3. The figures show average error rates (in %) ± standard deviations of the aggregated classifiers, measured from 10-fold crossvalidation for different α (solid line). The constant dashed line represents result of the unique, non-combined Bayesian classifier which uses all features.

In the case of the Phoneme dataset, the lowest average error rate was achieved by quasi-Sugeno system with $\wedge_{0.6}^{AA}$ – the improvement over non-combined classifier was about 4% (proved as significant improvement by the two sample t-test against the one-tailed alternative [achieved significance level was 0.00005]). For

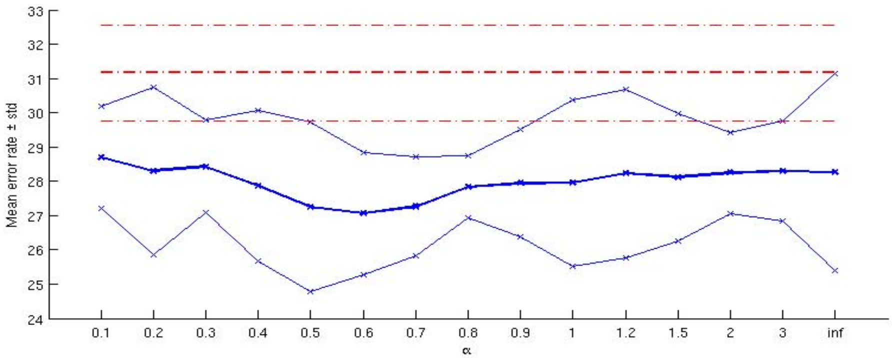

Fig. 2. Results for the Phoneme dataset. Solid line – ensemble aggregated using fuzzy t-conorm integral with $\mathcal{F} = (\vee_S, \vee_S, \vee_S, \wedge_\alpha^{AA})$, dashed line – non-combined classifier.

the Pima dataset, the improvement of about 1%, achieved by $\wedge_{0.7}^{AA}$, did not prove as statistically significant (significance level was 0.12).

What is even more important is that by using t-conorm integral (quasi-Sugeno with Aczel-Alsina t-norm in this case) we can fine-tune the parameters to obtain better results than those of the Sugeno integral. For both datasets, the improvement over Sugeno integral was about 1%, all that achieved without increasing the complexity of the algorithm. Although this improvement did not prove as statistically significant, achieved significance level was 0.15.

4.2 Experiment 2

A common drawback of many methods for combining classifiers is that if the data to classifiy are "easy", then the aggregated classifier often achieves worse results than a simple, non-combined classifier. In this experiment, we will show that for easy data (represented by the 4-dimensional Iris dataset from [15]), we can fine-tune the t-conorm system for integration, so that the performance of the aggregated classifier is only a slightly worse than the performance of non-combined classifier.

We created an ensemble of Bayesian classifiers using all possible subsets of features, resulting in 15 different classifiers. We aggregated the ensemble using quasi-Sugeno t-conorm integral with two different t-norms – Aczel-Alsina (which approaches $\wedge_D$ with $\alpha \to 0$), and Frank t-norm (which approaches $\wedge_L$ with $\alpha \to \infty$), defined as:

$$x \wedge_\alpha^F y = \begin{cases} x \wedge_S y & \text{if } \alpha = 0 \\ x \wedge_P y & \text{if } \alpha = 1 \\ x \wedge_L y & \text{if } \alpha = \infty \\ \log_\alpha \left(1 + \frac{(\alpha^x - 1)(\alpha^y - 1)}{\alpha - 1}\right) & \text{if } \alpha \in (0, \infty), \alpha \neq 1 \end{cases} \tag{21}$$

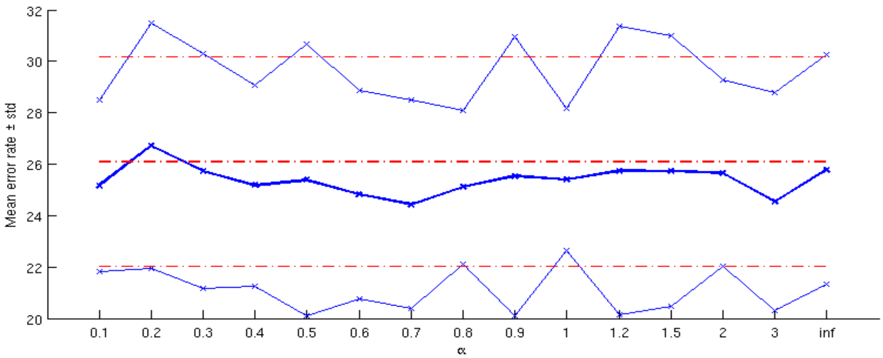

Fig. 3. Results for the Pima dataset. Solid line – ensemble aggregated using fuzzy t-conorm integral with $\mathcal{F} = (\vee_S, \vee_S, \vee_S, \wedge_\alpha^{AA})$, dashed line – non-combined classifier.

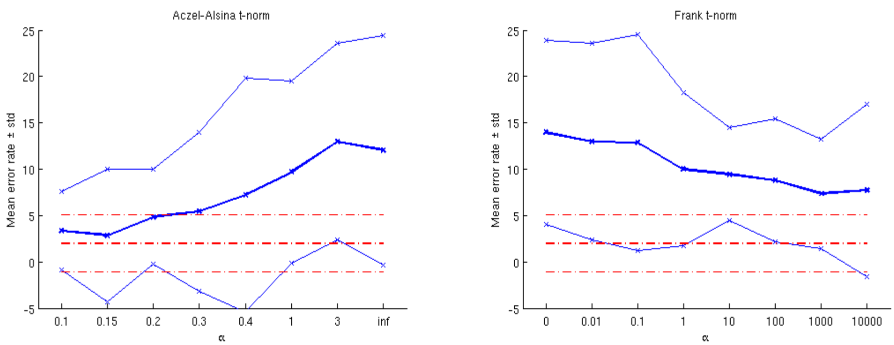

Fig. 4. Results for the Iris dataset. Solid line – ensemble aggregated using fuzzy t-conorm integral with $\mathcal{F} = (\vee_S, \vee_S, \vee_S, \wedge_\alpha^{AA})$ (left) or $\mathcal{F} = (\vee_S, \vee_S, \vee_S, \wedge_\alpha^{F})$ (right), dashed line – non-combined classifier.

In Fig. 4, the results for system with $\wedge_\alpha^{AA}$ for $\alpha = 0.1, 0.15, 0.2, 0.3, 0.4, 1, 3,$ and $\alpha = \infty$ (Sugeno integral), and $\wedge_\alpha^F$ for $\alpha = 0$ (Sugeno integral), $\alpha = 0.01, 0.1,$ $1, 10, 100, 1000, 10000$ are shown. We measured average error rates (in %) $\pm$ standard deviations from 10-fold crossvalidation (solid line); the dashed line corresponds to the unique, non-combined Bayesian classifier which uses all features.

The Iris dataset contains only about 150 patterns, so the results of the crossvalidation have big variance. We can see that Sugeno integral achieves average error rate about 12-14%. If we use quasi-Sugeno system with Frank t-norm, then as the t-norm approaches $\wedge_L$, the average error rate decreases ($\wedge_L$ cannot be used because it has zero divisors). Even better results were achieved by quasi-Sugeno system with Aczel-Alsina t-norm approaching $\wedge_D$.

5 Summary

In this paper, we described the fuzzy t-conorm integral and its use for combining classifiers. Different classes of t-conorm systems for integration were discussed. We showed that although the framework of fuzzy t-conorm integral is very general, only few t-conorm systems for integration can be used for combining classifiers (although the question which specific t-conorm system to use for a specific application remains unresolved). Still the fuzzy t-conorm integral adds additional degrees of freedom to classifier combining, and so it can provide more succesful way to classifier combining than Sugeno or Choquet fuzzy integral. That was confirmed also by tests on three benchmark datasets.

References

1. Murofushi, T., Sugeno, M.: Fuzzy t-conorm integral with respect to fuzzy measures: Generalization of Sugeno integral and Choquet integral. Fuzzy Sets and Systems 42(1), 57–71 (1991)
2. Melnik, O., Vardi, Y., Zhang, C.H.: Mixed group ranks: Preference and confidence in classifier combination. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(8), 973–981 (2004)
3. Kuncheva, L.I., Bezdek, J.C., Duin, R.P.W.: Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition 34(2), 299–314 (2001)
4. Breiman, L.: Bagging predictors. Machine Learning 24(2), 123–140 (1996)
5. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: International Conference on Machine Learning, pp. 148–156 (1996)
6. Bay, S.D.: Nearest neighbor classification from multiple feature subsets. Intelligent Data Analysis 3(3), 191–209 (1999)
7. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell. 20(3), 226–239 (1998)
8. Kuncheva, L.I.: Fuzzy versus nonfuzzy in combining classifiers designed by boosting. IEEE Transactions on Fuzzy Systems 11(6), 729–741 (2003)
9. Ahmadzadeh, M.R., Petrou, M.: Use of Dempster-Shafer theory to combine classifiers which use different class boundaries. Pattern Anal. Appl. 6(1), 41–46 (2003)
10. Grabisch, M., Nguyen, H.T.: Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference. Kluwer Academic Publishers, Norwell (1994)
11. Chiang, J.H.: Aggregating membership values by a Choquet-fuzzy-integral based operator. Fuzzy Sets Syst. 114(3), 367–375 (2000)
12. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer Academic Publishers, Dordrecht (2000)
13. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley, Chichester (2000)
14. Elena database: `http://www.dice.ucl.ac.be/mlg/?page=Elena`
15. Newman, D.J., Hettich, S., Merz, C.B.: UCI repository of machine learning databases (1998), `http://www.ics.uci.edu/~mlearn/MLRepository.html`