An Interval-Valued Fuzzy Classifier Based on an Uncertainty-Aware Similarity Measure

Anna Stachowiak, Patryk Żywica,
Krzysztof Dyczkowski, and Andrzej Wójtowicz

Faculty of Mathematics and Computer Science
Adam Mickiewicz University
Umultowska 87, 61-614 Poznań, Poland
{aniap, bikol, chris, andre}@amu.edu.pl

Abstract. In this paper we propose a new method for classifying uncertain data, modeled as interval-valued fuzzy sets. We develop the notion of an interval-valued prototype-based fuzzy classifier, with the idea of preserving full information including the uncertainty factor about data during the classification process. To this end, the classifier was based on the uncertainty-aware similarity measure, a new concept which we introduce and give an axiomatic definition for. Moreover, an algorithm for determining such a similarity value is proposed, and an application to supporting medical diagnosis is described.

1 Introduction

In this paper we present the novel concept of an interval-valued fuzzy classifier for supporting decision-making processes based on imprecise and incomplete (uncertain) data. The main aim was to develop a comprehensive, consistent and effective approach that enables the modeling and processing of input data, and then the presentation of results, to be done in a way that preserves the valuable information concerning the amount of uncertainty at each stage of the process.

Our theoretical framework is the well-developed interval-valued fuzzy set (IVFS) theory, introduced by Zadeh in [1] as a natural extension of fuzzy set theory. While a fuzzy set models a gradual, but precise, degree of truth of a statement, IVFSs make it possible to add uncertainty about that degree. IVFS $\widetilde{A}$ is defined by a pair of fuzzy sets $\underline{A}, \overline{A} : X \to [0,1]$ such that for each $x \in X$ an interval $[\underline{A}(x), \overline{A}(x)]$ is understood to contain the true, incompletely known membership degree. The length of this interval reflects the amount of uncertainty about an element x, taking values from 0 when $\underline{A}(x) = \overline{A}(x)$ to 1 when $[\underline{A}(x), \overline{A}(x)] = [0,1]$. Framework of IVFS was designed to be able to deal with incomplete data. In cases where data is missing, we assign the unit interval as the membership degree, meaning that all membership degrees are equally possible.

An IVFS is also a special case of type-2 fuzzy set (also introduced by Zadeh [1]), known by the name of interval type-2 fuzzy set. Another equivalent notion is the Atanassov intuitionistic fuzzy set theory, which specifies a membership function and a non-membership function separately (for a comparison, see [2]). All of these theories have proved their usefulness in many areas of soft computing such as fuzzy control, image processing, financial prediction, decision-making, computer vision, medical diagnosis, etc. [3, 4, 5, 6, 7]. In what follows we will contribute to medical decision support systems by constructing an interval-based fuzzy classifier.

Hereafter by $\mathcal{FS}$ we denote the set of all fuzzy sets, by $\mathcal{IVFS}$ we denote the set of all interval-valued fuzzy sets, and by $\mathcal{I}$ we denote the set of all intervals in $[0, 1]$. Formally a classifier can be defined as a function $D : \mathcal{O}^n \to \mathcal{C}$ where $\mathcal{O}$ is a set of possible values of classification features (it may be for example a subset of R or $\mathcal{I}$), $\mathcal{C}$ is the set of all possible m-element vectors which coordiantes correspond to particular classes and values reflect the membership to these classes, n is the number of classification features and m is the number of classes. As stated in [8], designing a classifier means "finding good D". D can be specified functionally or as a computer algorithm.

We can define different sets of labeled vectors in $\mathcal{C}$, resulting in different types of classifiers, e.g.:

1. crisp classifier:
$$\mathcal{C}_{cr} = \{y \in \{0, 1\}^m\}$$
2. probabilistic classifier:
$$\mathcal{C}_{pr} = \{y \in [0, 1]^m : \sum_{i=1}^{m} y_i = 1\}$$
3. fuzzy classifier:
$$\mathcal{C}_{fu} = \{y \in [0, 1]^m\}$$

There are many books on the theory and applications of fuzzy classification and pattern recognition. A very good introduction to designing fuzzy classifiers, with a very large subject bibliography, can be found in [8]; also worth examining are [9] and [10]. More information on the construction and use of crisp and probabilistic classifiers can be found in [11, 12].

In this article we introduce a new type of classifier:

4. interval-valued fuzzy classifier:
$$C_{iv} = \{y \in \mathcal{I}^m\}$$

A popular approach is to design, for each class, one prototype vector which represents the entire class. A very important element of such classification process is the construction of suitable vectors to describe the classes. Classifiers of this type are called prototype-based classifiers. Prototypes can be formed using clustering algorithms, such as k-means, or can result from the application of expert knowledge. We have taken the second approach, and through the use of IVFSs we were able to model experts knowledge along with its subjectivity, uncertainty and information deficiency.

In the next section we introduce in detail the idea of an uncertainty-aware similarity measure, which forms a central concept of a prototype-based classifier and enables the proper comparison of uncertain data. We propose an efficient algorithm for determining its value using the notion of relative cardinality of IVFSs. The third section is devoted to an interval-based classifier. In Section 4 we apply the ideas presented to the problem of supporting ovarian tumor diagnosis. Finally we state some conclusions.

2 Similarity measure for uncertain data

The similarity measure is a central concept of prototype-based classifiers that estimate the class label of a test sample based on the similarities between the test sample and a set of given prototypes.

In the following we briefly review some of the similarity measures for IVFSs known from the literature, and next we introduce the concept of an uncertainty-aware similarity measure, together with an effective algorithm for computing its value.

2.1 An overview of similarity measures for IVFSs

A common approach for measuring similarity involves the use of a distance metric. For example, in [13] the following similarity measure based on a normalized Hamming distance was proposed:

$$sim_H(\widetilde{A},\widetilde{B}) = 1 - \frac{1}{2n}\sum_{i=1}^{n}\left(|\underline{A}(x_i) - \underline{B}(x_i)| + \left|\overline{A}(x_i) - \overline{B}(x_i)\right|\right). \tag{1}$$

Another well-known approach is the Jaccard coefficient extended to IVFSs in the following way [14, 15]:

$$sim_J(\widetilde{A},\widetilde{B}) = \frac{\sum_{i=1}^{n}\min(\overline{A}(x_i),\overline{B}(x_i)) + \sum_{i=1}^{n}\min(\underline{A}(x_i),\underline{B}(x_i))}{\sum_{i=1}^{n}\max(\overline{A}(x_i),\overline{B}(x_i)) + \sum_{i=1}^{n}\max(\underline{A}(x_i),\underline{B}(x_i))} \tag{2}$$

Both sim_H and sim_J measure similarity with a single, scalar value. In the method proposed by Bustince [4] the similarity value is defined as an interval:

$$\begin{aligned}
sim_B(\widetilde{A},\widetilde{B}) &= \left[S_L(\widetilde{A},\widetilde{B}), S_U(\widetilde{A},\widetilde{B})\right] \\
S_L(\widetilde{A},\widetilde{B}) &= t\left(Inc_L(\widetilde{A},\widetilde{B}), Inc_L(\widetilde{B},\widetilde{A})\right) \\
S_U(\widetilde{A},\widetilde{B}) &= t\left(Inc_U(\widetilde{A},\widetilde{B}), Inc_U(\widetilde{B},\widetilde{A})\right) \\
Inc_L(\widetilde{A},\widetilde{B}) &= \inf_{x\in X}\left\{1, \min\left(1-\underline{A}(x)+\underline{B}(x), 1-\overline{A}(x)+\overline{B}(x)\right)\right\} \\
Inc_U(\widetilde{A},\widetilde{B}) &= \inf_{x\in X}\left\{1, \max\left(1-\underline{A}(x)+\underline{B}(x), 1-\overline{A}(x)+\overline{B}(x)\right)\right\}
\end{aligned} \tag{3}$$

where t is a t-norm i.e. an increasing, commutative and associative mapping $t : [0,1]^2 \to [0,1]$ satisfying $t(1,x) = x$ for all $x \in [0,1]$. More examples of IVFS similarity measures can be found in [14].

2.2 An uncertainty-aware similarity measure

The property of reflexivity of a similarity measure, natural and unquestionable in classical set theory, is not so obvious when comparing IVFSs with positive uncertainty value. Consider an extreme case of two different concepts represented by two one-element totally uncertain IVFS $\widetilde{A} = \widetilde{B} = {}^{[0,1]}/_x$. Mentioned IVFS, according to all classical similarity measures, including sim_H, sim_J and sim_B, are definitely identical ($sim_H(\widetilde{A}, \widetilde{B}) = sim_J(\widetilde{A}, \widetilde{B}) = 1$, $sim_B(\widetilde{A}, \widetilde{B}) = [1, 1]$). But is that also true for concepts they represent? In view of the total lack of knowledge, such a claim has no basis. In fact, the "real" membership degrees may be totally different for $\widetilde{A}$ and $\widetilde{B}$, and consequently a possible value of $sim(\widetilde{A}, \widetilde{B})$ may be anywhere in the range $[0, 1]$. Ignoring the uncertainty of $\widetilde{A}$ and $\widetilde{B}$ may have particularly adverse consequences when such a similarity measure is applied to practical problems. In this paper we emphasize the role and importance of the uncertainty factor for making well-informed decisions. For this reason we propose the notion of an uncertainty-aware similarity measure, described by the following definition.

Definition 1. *A mapping $sim_u : \mathcal{IVFS} \times \mathcal{IVFS} \to \mathcal{I}$ is said to be an uncertainty-aware similarity measure if it satisfies the following conditions for all $\widetilde{A}, \widetilde{B}, \widetilde{C}, \widetilde{D} \in \mathcal{IVFS}$:*

1. $sim_u(\widetilde{A}, \widetilde{B}) = [1, 1] \Leftrightarrow \widetilde{A} = \widetilde{B}$ & $\widetilde{A}, \widetilde{B} \in \mathcal{FS}$
2. $sim_u(\widetilde{A}, \widetilde{B}) = sim_u(\widetilde{B}, \widetilde{A})$
3. *if* $\widetilde{A} \subseteq \widetilde{B} \subseteq \widetilde{C}$ *then* $sim_u(\widetilde{A}, \widetilde{C}) \leq sim_u(\widetilde{A}, \widetilde{B})$ *and* $sim_u(\widetilde{A}, \widetilde{C}) \leq sim_u(\widetilde{B}, \widetilde{C})$
4. *if* $\widetilde{A} \sqsubseteq \widetilde{B}$ *and* $\widetilde{C} \sqsubseteq \widetilde{D}$ *then* $sim_u(\widetilde{A}, \widetilde{C}) \preceq sim_u(\widetilde{B}, \widetilde{D})$

where

- $[a, b] \leq [c, d] \Rightarrow a \leq c$ & $b \leq d$,
- $\widetilde{A} \subseteq \widetilde{B} \Leftrightarrow \underline{A}(x) \leq \underline{B}(x)$ & $\overline{A}(x) \leq \overline{B}(x), \forall x \in X$,
- $[a, b] \preceq [c, d] \Rightarrow a \geq c$ & $b \leq d$,
- $\widetilde{A} \sqsubseteq \widetilde{B} \Leftrightarrow \underline{A}(x) \geq \underline{B}(x)$ & $\overline{A}(x) \leq \overline{B}(x), \forall x \in X$.

Conditions 2 and 3 are standard symmetry and monotonicity properties. Conditions 1 and 4 reflect the goal of preserving the uncertainty value – when comparing uncertain objects, the final result should also be uncertain. Moreover, in 4 we require monotonicity with respect to uncertainty – the more certain are the compared IVFSs, the more certain is the resulting similarity measure sim_u.

In [15], and then in [16], [17] similarity for IVFSs was considered in the spirit of Definition 1. In the following we present an uncertainty-aware similarity measure of IVFSs based on the notion of relative cardinality of fuzzy sets, i.e. on the fact that for any $A, B \in \mathcal{FS}$:

$$sim(A, B) = \sigma(A \cap_t B | A \cup_t B) \tag{4}$$

where

$$\sigma(A|B) = \frac{\sigma(A \cap_t B)}{\sigma(B)}$$

is a relative cardinality of fuzzy sets. The scalar cardinality of fuzzy set $\sigma(A)$ is typically defined by the so-called sigma-count ([18, 19]):

$$\sigma(A) = \sum_{x \in X} A(x).$$

Moreover, sum $A \cup_t B$ and intersection $A \cap_t B$ are defined as:

$$\begin{aligned}(A \cup_t B)(x) &= t^* (A(x), B(x)) \\ (A \cap_t B)(x) &= t (A(x), B(x))\end{aligned}$$

where as a t-norm t we can take for example a minimum t-norm $t_\wedge(a,b) = a \wedge b$, a product t-norm $t_{\mathrm{prod}}(a,b) = a \cdot b$ or a Łukasiewicz t-norm $t_L(a,b) = 0 \vee (a+b-1)$. A t-conorm t^* is a dual operation such that $t^*(a,b) = 1 - t(1-a, 1-b)$.

To introduce a formula for an uncertainty-aware similarity measure, we first construct fuzzy representation sets of $\widetilde{A}$ and $\widetilde{B}$, where a representation set is defined as:

$$Rep(\widetilde{A}) = \{A \in \mathcal{FS} \mid \forall_{x \in X} \underline{A}(x) \le A(x) \le \overline{A}(x)\}.$$

Then the uncertainty-aware similarity measure sim_σ based on the notion of the relative cardinality, is defined in the following way.

Definition 2. *The cardinality-based uncertainty-aware similarity measure of two IVFSs $\widetilde{A} = (\underline{A}, \overline{A})$ and $\widetilde{B} = (\underline{B}, \overline{B})$, with a t-norm t, is an interval defined as:*

$$sim_\sigma(\widetilde{A}|\widetilde{B}) = \left[\underline{sim}_\sigma(\widetilde{A}|\widetilde{B}), \overline{sim}_\sigma(\widetilde{A}|\widetilde{B})\right] \tag{5}$$

where

$$\begin{aligned}\underline{sim}_\sigma(\widetilde{A}|\widetilde{B}) &= \min_{\substack{A \in Rep(\widetilde{A}) \\ B \in Rep(\widetilde{B})}} \sigma(A \cap_t B | A \cup_t B) \\ \overline{sim}_\sigma(\widetilde{A}|\widetilde{B}) &= \max_{\substack{A \in Rep(\widetilde{A}) \\ B \in Rep(\widetilde{B})}} \sigma(A \cap_t B | A \cup_t B)\end{aligned}$$

The formula given by Definition 2 is t-norm dependent. In the present paper we consider two of the most widely used t-norms: minimum and product. An effective algorithm for calculating a value of (5) for a minimum t-norm was given in [15]. In the following we present the algorithm for the case of the product t-norm that we introduced in [20].

The presented approach to measuring similarity is a key concept in the interval-valued classifier described in the next section. It allows the proper and effective comparison of imprecise and incomplete data, without losing information about the amount of uncertainty contained in that data.

Algorithm 1 Algorithms for computing lower (left) and upper (right) bounds of relative cardinality $\sigma_{IF}(\widetilde{A}|\widetilde{B})$ introduced in [20].

$n \leftarrow \sum_{x \in X} \underline{A}(x) \cdot \underline{B}(x)$
$d \leftarrow \sum_{x \in X} \underline{B}(x)$
for all $x \in X$ in descending order of $\underline{A}(x)$ **do**
    $r \leftarrow \frac{n}{d}$
    $n \leftarrow n + \underline{A}(x) \cdot \big(\overline{B}(x) - \underline{B}(x)\big)$
    $d \leftarrow d + \overline{B}(x) - \underline{B}(x)$
    if $r \geq \frac{n}{d}$ **then**
        return r
    end if
end for

$n \leftarrow \sum_{x \in X} \overline{A}(x) \cdot \underline{B}(x)$
$d \leftarrow \sum_{x \in X} \underline{B}(x)$
for all $x \in X$ in ascending order of $\overline{A}(x)$ **do**
    $r \leftarrow \frac{n}{d}$
    $n \leftarrow n + \overline{A}(x) \cdot \big(\overline{B}(x) - \underline{B}(x)\big)$
    $d \leftarrow d + \overline{B}(x) - \underline{B}(x)$
    if $r \leq \frac{n}{d}$ **then**
        return r
    end if
end for

3 Interval-valued fuzzy classifier

Our classifier is designed to deal with situations in which both the classified objects and the classes themselves are imprecise, subjective and/or incomplete. In such cases, the resulting classification would also be imprecise or incomplete. In our approach we take full account of these features of the data. The proposed classifier will be able to classify objects coded as an IVFS into classes which are also described in that way. Moreover, classification will also be described in interval-valued fashion. The final classification may be obtained with the use of the *score function* proposed in [21].

More formally, the problem can be formulated in the following way. Given a set $\mathcal{O}$ of objects to classify, described as an IVFS, compute for each of them the IVFS in the domain of set $\mathcal{C}$ of all possible classes and its interval-valued membership which describe the degree to which the object belongs to given class.

In this paper we assume that class prototypes as well as objects are coded as IVFSs. For instance class $c \in \mathcal{C}$ is coded as IVFS $\widetilde{iv}(c)$, and object $o_i \in \mathcal{O}$ as $\widetilde{iv}(o_i)$. The intuition behind the proposed classifier is to use an uncertainty-aware similarity measure to compute the similarities between objects and class prototypes and then use them as membership degrees in the resulting classification. Formally, the assignment of object $o_i \in \mathcal{O}$ to classes using the singleton notation can be stated as follows:

$$\widetilde{A}_{o_i} = \sum_{c \in \mathcal{C}} {}^{sim_{IF}(\widetilde{iv}(c), \widetilde{iv}(o_i))}\!/_{c}$$

It should be noted that uncertainty-aware similarity plays a fundamental role in our method. Thus it is crucial to use a similarity measure applicable to the problem being solved. $sim_{IF}(\widetilde{iv}(c), \widetilde{iv}(o_i)$ is an interval, hence the resulting classification $\widetilde{A}_{o_i}$ is an IVFS.

4 Application to supporting ovarian tumor diagnosis

The proposed classifier is demonstrated on a real data set from the field of medicine. Ovarian tumors are currently one of the most deadly diseases among women. According to recent statistics, annual numbers of new cases and deaths in the USA amount to 22,000 and 14,000 [22].

Two main groups of tumors are discriminated: malignant and benign. Such a division turns the diagnostic process into a classic binary classification. Each of these groups subdivides into histopathological types. This means that the problem might be expanded to a multi-class classification.

The correct discrimination between ovarian tumors is a key issue, because it determines the method of treatment. For this reason, a number of preoperative models for malignancy discrimination have been developed over the past two decades. They range from basic scoring systems [23, 24] to formal mathematical models, in particular rule-based schemes [25] and machine learning techniques [26, 27]. Unfortunately, in external evaluation the efficacy of predictions of such models rarely exceeds 90%, in terms of both sensitivity and specificity [28, 29]. Therefore, there is still a need to develop an effective preoperative model for inexperienced gynecologists.

In our previous research we have indicated the imprecision of data obtained by a gynecologist during examinations [30]. Many features are undoubtedly objective, such as levels of blood markers. Some examinations, however, require assessment on the part of the gynecologist, who can be a source of subjectivity – the more experienced he/she is, the more confident the result. In particular, this is the case when an ultrasonographer examines a tumor. Furthermore, in some cases we also have to deal with lack of data. Some examinations might be omitted by a physician, either for medical reasons or due to the their unavailability. These attributes may be conveniently modeled using IVFS theory.

The interval-valued classifier described above was applied to the problem of supporting ovarian tumor diagnosis. In this problem we try to assign the best matching histopathological profile of a tumor using the data available before an operation.

Both patient and histopathological profiles were coded as IVFSs. For the sake of simplicity, only four histopathological types were modeled for the present example. Two of them were benign – *Endometrioid cyst* and *Mucinous cystadenoma* – and two malignant – *Serous adenocarcinoma* and *Undifferentiated carcinoma* – referred to further as HP 1, HP 6, HP 21 and HP 25 respectively. Characteristics of class prototypes were obtained from an experts knowledge and partially from analysis of historical medical data. Among more than fifty attributes describing patient, five were arbitrarily selected: age, size of papillary projections (PAP), blood serum levels of CA-125 and HE4 tumor markers, and resistive index (RI). These attributes may be more or less subjective or imprecise. Moreover, some data may be not available at all. Based on a survey among gynecologists at Poznań University of Medical Sciences, characteristics of those parameters were obtained. A patients age is known precisely, while blood serum levels of tumor markers are subject to some uncertainties. Resistive index and size of papillary projections were indicated as subjective attributes in the survey, thus their value is uncertain. Moreover, values of the last three attributes were not always available.

Real data based on histopathological profiles and example patients are presented in Tables 1 and 2. Note that patients missing data were replaced with the unit interval $[0, 1]$.

HP type	AGE	PAP	CA125	HE4	RI
1	[0.27, 0.64]	[0, 0.27]	[0, 0.04]	[0, 0.03]	[0.49, 0.78]
6	[0.29, 0.72]	[0, 0.14]	[0, 0.18]	[0.01, 0.06]	[0.22, 0.83]
21	[0.47, 0.76]	[0, 0.52]	[0.3, 1]	[0.12, 0.9]	[0.23, 0.56]
25	[0.39, 0.77]	[0, 0.58]	[0.15, 0.98]	[0.04, 0.62]	[0.27, 0.45]

Table 1. Profiles of ovarian tumor histopathological type coded as IVFS.

Patient #	Postoperative diagnosis	AGE	PAP	CA125	HE4	RI
1	HP 1	[0.78, 0.78]	[0.00, 0.38]	[0.00, 0.06]	[0.00, 1.00]	[0.00, 1.00]
2	HP 2	[0.14, 0.14]	[0.35, 0.85]	[0.05, 0.15]	[0.00, 1.00]	[0.60, 1.00]
3	HP 21	[0.62, 0.62]	[0.00, 0.25]	[0.95, 1.00]	[0.95, 1.00]	[0.00, 1.00]

Table 2. Profiles of patients coded as IVFS.

Now a classification using the cardinality-based uncertainty-aware similarity measure formulated by (5) can be computed. By definition, the third patients classification is the following:

$$\begin{aligned}\widetilde{A}_{o_3} = &\, {}^{sim_u(\widetilde{iv}(o_3),\widetilde{iv}(hp_3))}/_{hp_3} + {}^{sim_u(\widetilde{iv}(o_3),\widetilde{iv}(hp_6))}/_{hp_6} \\ &+ {}^{sim_u(\widetilde{iv}(o_3),\widetilde{iv}(hp_{21}))}/_{hp_{21}} + {}^{sim_u(\widetilde{iv}(o_3),\widetilde{iv}(hp_{25}))}/_{hp_{25}}\end{aligned}$$

If we use the minimum t-norm for calculation of similarity, the final classification is as follows:

$$\begin{aligned}\widetilde{A}_{o_3} = &\, {}^{[0.07,0.48]}/_{hp_1} + {}^{[0.08,0.52]}/_{hp_6} \\ &+ {}^{[0.21,0.99]}/_{hp_{21}} + {}^{[0.13,0.90]}/_{hp_{25}}\end{aligned} \tag{6}$$

and in the case of the product t-norm:

$$\begin{aligned}\widetilde{A}_{o_3} = &\, {}^{[0.50,0.98]}/_{hp_1} + {}^{[0.31,0.98]}/_{hp_6} \\ &+ {}^{[0.54,0.99]}/_{hp_{21}} + {}^{[0.47,0.99]}/_{hp_{25}}\,.\end{aligned} \tag{7}$$

The interpretation of classification from (6) is as follows. The possibility that this patient should be diagnosed as HP 1 is [0.07, 0.48], which is a low score. By contrast, the membership of class HP 21 is [0.21, 0.99]. To choose the best matching class we can use the score function defined in [21] as

$$\text{score}([a, b]) = a + b - 1 .$$

Using this score, it is easy to see that class HP 21 is the best match for this patient (with a score of 0.20, compared with -0.45, -0.4 and 0.03). It is worth noting that this patient was postoperatively diagnosed with a type HP 21 tumor.

Tables 3 and 4 contain both classifications for the patients represented by interval and score result.

Patient #	HP1		HP6		HP21		HP25	
	interval	score	interval	score	interval	score	interval	score
1	[0.08, 0.93]	0.01	[0.09, 0.97]	0.06	[0.11, 0.91]	0.02	[0.09, 0.96]	0.05
2	[0.17, 0.85]	0.02	[0.10, 0.79]	-0.11	[0.09, 0.81]	-0.10	[0.10, 0.83]	-0.07
3	[0.07, 0.48]	-0.45	[0.08, 0.52]	-0.4	[0.21, 0.99]	0.20	[0.13, 0.90]	0.03

Table 3. Classification obtained using minimum t-norm.

Patient #	HP1		HP6		HP21		HP25	
	interval	score	interval	score	interval	score	interval	score
1	[0.50, 0.98]	0.48	[0.32, 0.99]	0.31	[0.28, 0.99]	0.27	[0.25, 0.99]	0.24
2	[0.54, 0.99]	0.53	[0.40, 0.99]	0.39	[0.22, 0.99]	0.21	[0.20, 0.99]	0.19
3	[0.50, 0.98]	0.48	[0.31, 0.98]	0.29	[0.54, 0.99]	0.53	[0.47, 0.99]	0.48

Table 4. Classification obtained using product t-norm.

5 Conclusions

The main contribution of this paper is the concept of an uncertainty-aware similarity measure, which was axiomatically defined and for which an efficient algorithm was given. Based on this measure an interval-valued fuzzy classifier was constructed and applied to real-life data concerning patients with ovarian tumors. The uncertainty-aware similarity measure proved to be particularly useful for supporting medical diagnosis, where uncertainty and incompleteness of information are common and inevitable features. We obtained promising results showing that IVFS theory is convenient for modeling and processing such data, and our experience suggests that practitioners prefer informative, even if uncertain, feedback rather than excessively precise data. In our further research we plan to improve the classification process and to thoroughly investigate the properties of the uncertainty-aware similarity measure, including the influence of the t-norm used in the similarity formula.

References

[1] Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning–I. Information Sciences **8**(3) (1975) 199–249

[2] Atanassov, K.T.: Intuitionistic fuzzy sets. Fuzzy Sets and Systems **20**(1) (1986) 87–96

[3] Deschrijver, G., Kerre, E.: Advances and challenges in interval-valued fuzzy logic. Fuzzy Sets and Systems **157** (2006) 622–627

[4] Bustince, H.: Indicator of inclusion grade for interval valued fuzzy sets. Application to approximate reasoning based on interval-valued fuzzy sets. International Journal of Approximate Reasoning **23** (2000) 137–209

[5] Wu, D.: On the fundamental differences between interval type-2 and type-1 fuzzy logic controllers. IEEE Transactions on Fuzzy Systems **20** (2012) 832–848

[6] Yager, R.R.: Fuzzy subsets of type II in decisions. Journal of Cybernetics **10** (1980) 137–159

[7] Karnik, N.N., Mendel, J.M.: Applications of type-2 fuzzy logic systems to forecasting of time-series. Information Sciences **120** (1999) 89–111

[8] Bezdek, J., Keller, J., Krisnapuram, R., Pal, N.: Fuzzy Models and algorithms for pattern recognition and image processing. Springer (2005)

[9] Kuncheva, L.: Fuzzy Classifier Design. Studies in Fuzzines and Soft Computing. Springer (2000)

[10] Höppner, F., Klawonn, F., Kruse, R., Runkler, T.: Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition. Wiley (1999)

[11] Duda, R., Hart, P., Stork, D.: Pattern Classification, 2nd Edition. Willey (2000)

[12] Bishop, C.: Pattern Recognition and Machine Learning. Information Science and Statistics. Springer (2006)

[13] Zeng, W., Li, H.: Relationship between similarity measure and entropy of interval valued fuzzy sets. Fuzzy Sets and Systems **157** (2006) 1477–1484

[14] Wu, D., Mendel, J.M.: A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets. Information Sciences **179** (2009) 1169–1192

[15] Nguyen, H.T., Kreinovich, V.: Computing degrees of subsethood and similarity for interval-valued fuzzy sets: Fast algorithms. In: Proc. of the 9th International Conference on Intelligent Technologies InTec08. (2008) 47–55

[16] Wu, D., Mendel, J.: Efficient algorithms for computing a class of subsethood and similarity measures for interval type-2 fuzzy sets. In: FUZZ-IEEE. (2010) 1–7

[17] Stachowiak, A., Dyczkowski, K.: A similarity measure with uncertainty for incompletely known fuzzy sets. In: Proceedings of the 2013 Joint IFSA World Congress NAFIPS Annual Meeting. (2013) 390–394

[18] Luca, A.D., Termini, S.: A definition of nonprobabilistic entropy in the setting of fuzzy sets theory. Information and Computation **20** (1972) 301–312

[19] Szmidt, E., Kacprzyk, J.: Entropy for intuitionistic fuzzy sets. Fuzzy Sets and Systems **118** (2001) 467–477

[20] Żywica, P., Stachowiak, A.: A new method for computing relative cardinality of intuitionistic fuzzy sets. In Krassimir Atanassov, e.a., ed.: New Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. IBS PAN – SRI PAS, Warsaw (2014) to appear.

[21] Chen, S., Tan, J.M.: Handling multicriteria fuzzy decision-making problems based on vague set theory. Fuzzy Sets and Systems **67** (1994) 163–172

[22] Siegel, R., Ma, J., Zou, Z., Jemal, A.: Cancer statistics, 2014. CA: A Cancer Journal for Clinicians **64**(1) (2014) 9–29

[23] Alcázar, J.L., Mercé, L.T., Laparte, C., Jurado, M., López-Garcia, G.: A new scoring system to differentiate benign from malignant adnexal masses. Obstetrical & Gynecological Survey **58**(7) (2003) 462–463

[24] Szpurek, D., Moszyński, R., Ziętkowiak, W., Spaczyński, M., Sajdak, S.: An ultrasonographic morphological index for prediction of ovarian tumor malignancy. European Journal of Gynaecological Oncology **26**(1) (2005) 51–54

[25] Timmerman, D., Testa, A.C., Bourne, T., Ameye, L., Jurkovic, D., Van Holsbeke, C., Paladini, D., Van Calster, B., Vergote, I., Van Huffel, S., et al.: Simple ultrasound-based rules for the diagnosis of ovarian cancer. Ultrasound in Obstetrics & Gynecology **31**(6) (2008) 681–690

[26] Timmerman, D., Testa, A.C., Bourne, T., Ferrazzi, E., Ameye, L., Konstantinovic, M.L., Van Calster, B., Collins, W.P., Vergote, I., Van Huffel, S., et al.: Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group. Journal of Clinical Oncology **23**(34) (2005) 8794–8801

[27] Van Calster, B., Timmerman, D., Nabney, I.T., Valentin, L., Testa, A.C., Van Holsbeke, C., Vergote, I., Van Huffel, S.: Using Bayesian neural networks with ARD input selection to detect malignant ovarian masses prior to surgery. Neural Computing and Applications **17**(5-6) (2008) 489–500

[28] Valentin, L., Hagen, B., Tingulstad, S., Eik-Nes, S.: Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. Ultrasound in Obstetrics & Gynecology **18**(4) (2001) 357–365

[29] Van Holsbeke, C., Van Calster, B., Valentin, L., Testa, A.C., Ferrazzi, E., Dimou, I., Lu, C., Moerman, P., Van Huffel, S., Vergote, I., et al.: External validation of mathematical models to distinguish between benign and malignant adnexal tumors: a multicenter study by the International Ovarian Tumor Analysis Group. Clinical Cancer Research **13**(15) (2007) 4440–4447

[30] Wójtowicz, A., Żywica, P., Szarzyński, K., Moszyński, R., Szubert, S., Dyczkowski, K., Stachowiak, A., Szpurek, D., Wygralak, M.: Dealing with Uncertainty in Ovarian Tumor Diagnosis. In Krassimir Atanassov, e.a., ed.: New Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. IBS PAN – SRI PAS, Warsaw (2014) to appear.