A First Study on the Use of Interval-Valued Fuzzy Sets with Genetic Tuning for Classification with Imbalanced Data-Sets

J. Sanz[1,⋆], A. Fernández[2], H. Bustince[1], and F. Herrera[2]

[1] Dept. of Automatic and Computation, Public University of Navarra, Spain
Tel.:+34-948169839; Fax:+34-948168924
{joseantonio.sanz,bustince}@unavarra.es
[2] Dept. of Computer Science and A.I., University of Granada
{alberto,herrera}@decsai.ugr.es

Abstract. Classification with imbalanced data-sets is one of the recent challenging problems in Data Mining. In this framework, the class distribution is not uniform and the separability between the classes is often difficult. From the available techniques in the Machine Learning field, we focus on the use of Fuzzy Rule Based Classification Systems, as they provide an interpretable model for the end user by means of linguistic variables.

The aim of this work is to increase the performance of fuzzy modeling by adding a higher degree of knowledge by means of the use of Interval-valued Fuzzy Sets. Furthermore, we will contextualize the Interval-valued Fuzzy Sets with a post-processing genetic tuning of the amplitude of their upper bounds in order to enhance the global behaviour of this methodology.

Keywords: Fuzzy Rule-Based Classification Systems, Interval-valued Fuzzy Sets, Tuning, Genetic Algorithms, Imbalanced Data-Sets.

1 Introduction

When facing a classification problem, the user can choose among many techniques to solve it. One of them, known as Fuzzy Rule-Based Classification Systems (FRBCS)[1], is mostly employed because of its interpretability and the possibility of mixing different kinds of information as the one given by experts and the one that comes from mathematical models or empiric measures.

In this work, we will deal with one of the emergent challenging problems in Data Mining [2], the classification with imbalanced data-sets [3]. Specifically, we will focus in the two-class imbalanced problem which appears when one class (known as positive class) is represented by only a few examples, whereas the other (negative class) is described by many instances. Furthermore, it is common that the positive class is the most interesting one from the point of view of the learning task. We can find some recent works in the literature that study the effect of imbalance between the classes in the framework of FRBCSs [4].

⋆ Corresponding author.

E. Corchado et al. (Eds.): HAIS 2009, LNAI 5572, pp. 581–588, 2009.

Standard classifier algorithms tend to be biased towards the negative class, since the rules that predict the highest number of examples are rewarded by the accuracy metric. Our aim here is to improve the performance of FRBCSs using the model of Interval-valued Fuzzy Sets (IVFSs) [5]. Specifically, we consider that the success of the use of fuzzy set theory depends on the choice of the membership function (MF) but, when experts do not have precise knowledge of the function to be taken, or it is defined ad-hoc, it can be appropriate to represent the membership degree of each element by means of an interval. Hence, not only vagueness (lack of sharp class boundaries) but also a feature of uncertainty (lack of information) can be addressed intuitively.

We will apply a post-processing step for tuning the amplitude of the upper bounds in the IVFSs, contextualizing the fuzzy partitions for the problem to solve. This is necessary because the data distribution is not necessary uniform and the amplitude of each label may be different.

To build the initial Knowledge Base (KB) we will employ the Chi et al.'s method [6] and we will compare the IVFS methodology (with and without tuning) against the results obtained with this initial KB. Furthermore, we will include the C4.5 decision tree in our experimental study, since it is an algorithm of reference in the field of imbalanced data-sets [7,8]. To do so, we will employ forty four data-sets from UCI repository [9], where multi-class data sets are modified to obtain two-class non-balanced problems, defining the joint of one or more classes as positive and the joint of one or more classes as negative. To evaluate our results we have applied the Area Under the Curve (AUC) metric [10] carrying out some non-parametric tests [11,12] to show the significance in the performance improvements obtained.

This work is organized as follows: in Section 2 we describe the problem of imbalanced data-sets. In Section 3 we define the IVFS model. Section 4 introduces our experimentation framework and shows the experimental study. In Section 5 we summarize the study carried out.

2 Imbalanced Data-Sets in Classification

The problem of imbalanced data-sets in classification [3] occurs when the class distribution is not uniform. In this situation, the number of examples that represents one class of the data-set (usually the concept of interest) is much lower than that of the other class. This situation has been recently identified as one important problem in data mining, since it is implicit in most real applications including telecommunications, finances, biology or medicine.

This scenario may suppose an added difficulty for the identification and discovery of rules covering the under-represented samples. In [4], the authors studied different configurations for FRBCSs in order to determine the most suitable model in this classification framework. Furthermore, it is shown the necessity to apply a re-sampling procedure; specifically, the "Synthetic Minority Oversampling Technique" (SMOTE) [13] obtains a very good behaviour.

As we stated before, most of proposals for automatic learning of classifiers use some kind of accuracy measure as the classification percentage over the example set. However, these measures can lead to erroneous conclusions over imbalanced data-sets since they don't take into account the proportion of examples for each class. Therefore, in this work we use the AUC metric [10], defined as

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2}, \tag{1}$$

where TP_{rate} and FP_{rate} are the percentage of correctly and wrongly classified cases belonging to the positive class respectively.

3 Interval-Valued Fuzzy Sets and Amplitude Tuning

In this work we want to improve the performance of FRBCSs applying IVFS to represent the different fuzzy partitions. We will use the Chi et al.'s rule learning algorithm [6], where we represent fuzzy rules as:

$$\text{Rule } R_j : \text{ If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \text{ then Class} = C_j \text{ with } RW_j, \tag{2}$$

where R_j is the label of the jth rule, $x = (x_1, \ldots, x_n)$ is an n-dimensional pattern vector, A_{ji} is an antecedent fuzzy set, C_j is a class label, and RW_j is the rule weight. We represent the MFs by triangular functions.

In the remaining of this section, we will first describe the IVFSs model and then we will present the genetic tuning of the amplitude for the fuzzy labels.

3.1 IVFSs Model

The IVFSs [5] are an extension of the theory of fuzzy sets which enables to manage additional knowledge in the fuzzy partitions. In the following we define this model with some detail:

We denote by $L([0,1])$ the set of all closed subintervals of the closed interval $[0,1]$; that is: $L([0,1]) = \{\mathbf{x} = [\underline{x}, \overline{x}] | (\underline{x}, \overline{x}) \in [0,1]^2 \text{ and } \underline{x} \leq \overline{x}\}$.
$L([0,1])$ is a partially ordered set with respect to the relation $\leq_L$ defined in the following way; given $\mathbf{x}, \mathbf{y} \in L([0,1])$: $\mathbf{x} \leq_L \mathbf{y}$ if and only if $\underline{x} \leq \underline{y}$ and $\overline{x} \leq \overline{y}$.
$(L([0,1]), \leq_L)$ is a complete lattice where the smallest element is $0_L = [0,0]$ and the largest is $1_L = [1,1]$.

Definition 1. *An Interval-valued fuzzy set (IVFS) A on the universe $U \neq \emptyset$ is a mapping $A : U \to L([0,1])$.*

Obviously, $A(u) = [\underline{A}(u), \overline{A}(u)] \in L([0,1])$ is the membership degree of u $\in U$.

We generate the initial KB following a simple rule learning algorithm (Chi el al.'s method in this case) and, from this KB, we include the IVFSs model by adding an upper bound for each fuzzy partition, centered in the maximum of the MF and with a higher amplitude. In our initial model, the amplitude of the upper bound will be 50% greater than that of the lower bound. We must remark that we note "upper" and "lower" bounds referring to the corresponding fuzzy labels.

Now, we are working with an interval when computing the matching degree between the antecedent of the rule and the example. In order to give a single output value, we obtain the mean between the lower and the upper matching degrees. Furthermore, the rule weight is composed by two numbers, associated to the lower and the upper bound respectively, and the same procedure will be employed in this case. Specifically, the rule weight is computed using the Penalized Certainty Factor defined in [14] as:

$$CF_{Lj} = \frac{\sum\limits_{x_p \in ClassC_j} \underline{A_j}(x_p) - \sum\limits_{x_p \notin ClassC_j} \underline{A_j}(x_p)}{\sum\limits_{p=1}^{m} \underline{A_j}(x_p)} \tag{3}$$

Note that for the rule weight computation of the upper bound, we may only replace $\underline{A_j}(x_p)$ with $\overline{A_j}(x_p)$.

3.2 Genetic Tuning of the Amplitude of Upper Bound of the IVFS

To improve the performance of the initial IVFSs model, we have to contextualize the fuzzy partitions for each problem. To do so, we propose a genetic tuning approach to perform slight changes of the original upper bound amplitude.

The modification of the amplitude is given by a number within the interval [0, 1], that is, from the overlapping of both bounds (value 0) to twice the amplitude of the upper with respect to the lower bound (value 1). The amplitude of the upper bound will be uniformly increased according to intermediate values.

In order to apply the genetic tuning, we will consider the use of CHC algorithm [15], which presents a good trade-off between diversity and convergence, being a good choice in complex problems. The components needed to design this process are explained below:

1. *Coding Scheme:* A real coding is considered, where each gene of the chromosome represents the amplitude modifier as defined above. Thus, there are as many genes as fuzzy partitions in the Data Base.
2. *Chromosome Evaluation:* The fitness function is the AUC metric.
3. *Initial Gene Pool:* The initial pool is obtained with the first individual having all genes with value '0.5' (the initial FRBCS). The second and the third individuals having all genes with values 0 and 1 respectively, whereas the remaining individuals are generated at random in [0, 1].
4. *Crossover Operator:* We consider the Parent Centric BLX (PCBLX) operator, which is based on the BLX-α. We consider the incest prevention mechanism, checking and modifying an initial threshold, in order to apply the PCBLX operator.
5. *Restarting approach:* When the threshold value is lower than zero, all the chromosomes are regenerated at random within the interval [0, 1]. Furthermore, the best global solution found is included in the population to increase the convergence of the algorithm.

4 Experimental Study

In this study, our intention is to show the improvement achieved in FRBCSs applying the IVFSs model. To do this we have to do a double analysis:

– We want to analyze whether the IVFSs model enhances the performance of a simple KB.
– We want to determine the significance of the tuning step in the IVFSs model.

In the remaining of this section, we will first present the experimental framework and all the parameters employed in this study and then we will show the empirical study for the IVFSs model in imbalanced data-sets.

4.1 Experimental Set-Up

To carry out the different experiments we consider a *5-folder cross-validation model*, i.e., 5 random partitions of data with a 20%, and the combination of 4 of them (80%) as training and the remaining one as test. For each data-set we consider the average results of the five partitions. Furthermore, a Wilcoxon's Signed-Ranks Test [16] is used for statistical comparison of our empirical results. In all cases the level of confidence (α) will be set at 0.05.

Table 1. Summary Description for Imbalanced Data-Sets

Data-set	#Ex.	#Atts.	Class (min., maj.)	%Class(min.; maj.)
Glass1	214	9	(build-win-non_float-proc; remainder)	(35.51, 64.49)
Ecoli0vs1	220	7	(im; cp)	(35.00, 65.00)
Wisconsin	683	9	(malignant; benign)	(35.00, 65.00)
Pima	768	8	(tested-positive; tested-negative)	(34.84, 66.16)
Iris0	150	4	(Iris-Setosa; remainder)	(33.33, 66.67)
Glass0	214	9	(build-win-float-proc; remainder)	(32.71, 67.29)
Yeast1	1484	8	(nuc; remainder)	(28.91, 71.09)
Vehicle1	846	18	(Saab; remainder)	(28.37, 71.63)
Vehicle2	846	18	(Bus; remainder)	(28.37, 71.63)
Vehicle3	846	18	(Opel; remainder)	(28.37, 71.63)
Haberman	306	3	(Die; Survive)	(27.42, 73.58)
Glass0123vs456	214	9	(non-window glass; remainder)	(23.83, 76.17)
Vehicle0	846	18	(Van; remainder)	(23.64, 76.36)
Ecoli1	336	7	(im; remainder)	(22.92, 77.08)
New-thyroid2	215	5	(hypo; remainder)	(16.89, 83.11)
New-thyroid1	215	5	(hyper; remainder)	(16.28, 83.72)
Ecoli2	336	7	(pp; remainder)	(15.48, 84.52)
Segment0	2308	19	(brickface; remainder)	(14.26, 85.74)
Glass6	214	9	(headlamps; remainder)	(13.55, 86.45)
Yeast3	1484	8	(me3; remainder)	(10.98, 89.02)
Ecoli3	336	7	(imU; remainder)	(10.88, 89.12)
Page-blocks0	5472	10	(remainder; text)	(10.23, 89.77)
Yeast2vs4	514	8	(cyt; me2)	(9.92, 90.08)
Yeast05679vs4	528	8	(me2; mit,me3,exc,vac,erl)	(9.66, 90.34)
Vowel0	988	13	(hid; remainder)	(9.01, 90.99)
Glass016vs2	192	9	(ve-win-float-proc; build-win-float-proc, build-win-non_float-proc,headlamps)	(8.89, 91.11)
Glass2	214	9	(Ve-win-float-proc; remainder)	(8.78, 91.22)
Ecoli4	336	7	(om; remainder)	(6.74, 93.26)
Yeast1vs7	459	8	(nuc; vac)	(6.72, 93.28)
Shuttle0vs4	1829	9	(Rad Flow; Bypass)	(6.72, 93.28)
Glass4	214	9	(containers; remainder)	(6.07, 93.93)
Page-blocks13vs2	472	10	(graphic; horiz.line,picture)	(5.93, 94.07)
Abalone9vs18	731	8	(18; 9)	(5.65, 94.25)
Glass016vs5	184	9	(tableware; build-win-float-proc, build-win-non_float-proc,headlamps)	(4.89, 95.11)
Shuttle2vs4	129	9	(Fpv Open; Bypass)	(4.65, 95.35)
Yeast1458vs7	693	8	(vac; nuc,me2,me3,pox)	(4.33, 95.67)
Glass5	214	9	(tableware; remainder)	(4.20, 95.80)
Yeast2vs8	482	8	(pox; cyt)	(4.15, 95.85)
Yeast4	1484	8	(me2; remainder)	(3.43, 96.57)
Yeast1289vs7	947	8	(vac; nuc,cyt,pox,erl)	(3.17, 96.83)
Yeast5	1484	8	(me1; remainder)	(2.96, 97.04)
Ecoli0137vs26	281	7	(pp,imL; cp,im,imU,imS)	(2.49, 97.51)
Yeast6	1484	8	(exc; remainder)	(2.49, 97.51)
Abalone19	4174	8	(19; remainder)	(0.77, 99.23)

Table 2. Results for FRBCSs and C4.5 in imbalanced data-sets. By columns we represent the Chi et al.'s algorithm with the lower bound (Chi_Low), Chi with the upper bound (Chi_Up), IVFSs model (Chi_IVFS), IVFS with tuning (Chi_IVFS_tun) and C4.5.

Data-set	Chi_Low		Chi_Up		Chi_IVFS		Chi_IVFS_tun		C4.5	
	AUC_{Tr}	AUC_{Tst}	AUC_{Tr}	AUC_{Tst}	AUC_{Tr}	AUC_{Tst}	AUC_{Tr}	AUC_{Tst}	AUC_{Tr}	AUC_{Tst}
Glass1	75.54	65.53	69.28	67.76	72.06	66.31	80.37	71.75	89.78	75.77
Ecoli0vs1	95.61	92.71	98.05	96.04	96.42	94.04	98.54	95.38	99.27	97.96
Wisconsin	98.07	89.19	97.08	96.03	97.32	96.03	98.24	96.43	98.32	95.45
Pima	72.64	67.66	67.23	65.93	70.03	67.69	76.04	70.80	84.11	71.45
Iris0	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.00
Glass0	71.68	69.74	72.18	72.18	72.01	71.83	73.92	71.81	94.33	78.56
Yeast1	70.12	69.44	67.57	67.81	68.55	68.71	72.96	70.97	80.49	71.09
Vehicle1	76.86	71.40	71.71	69.34	72.55	69.26	78.87	71.80	95.51	70.30
Vehicle2	88.13	85.55	83.30	81.45	83.83	82.29	93.56	88.28	98.95	94.92
Vehicle3	75.96	69.51	71.66	67.28	71.72	67.19	77.49	69.34	94.93	74.44
Haberman	66.98	60.60	65.79	55.06	67.13	59.89	70.75	61.06	74.26	63.09
Glass0123vs456	94.08	86.42	93.18	90.09	94.13	92.02	96.29	92.33	99.08	90.32
Vehicle0	88.75	86.96	80.82	80.07	82.27	81.62	90.98	87.13	98.97	91.18
Ecoli1	87.95	85.88	90.35	88.51	90.31	87.91	91.84	85.72	96.31	77.55
New-Thyroid2	94.80	90.60	96.73	96.31	97.15	90.87	99.51	98.02	99.57	96.59
New-Thyroid1	92.60	88.33	97.73	96.31	95.16	90.04	99.30	96.31	99.22	98.02
Ecoli2	89.68	88.26	89.09	87.68	89.14	88.34	91.54	87.44	95.17	91.62
Segment0	95.53	95.07	86.39	86.31	91.37	91.38	97.69	96.64	99.85	99.27
Glass6	95.07	84.69	92.14	84.74	93.99	85.01	97.33	84.42	99.59	84.50
Yeast3	91.43	90.22	85.35	84.55	88.21	86.69	93.15	91.21	95.65	88.76
Ecoli3	89.38	87.84	88.17	86.60	88.19	86.95	91.98	90.30	98.15	89.21
Page-Blocks0	81.89	81.40	80.34	80.32	81.22	80.95	84.35	83.68	98.46	94.85
Yeast2vs4	89.68	87.36	89.39	88.28	90.12	88.25	91.04	88.15	98.14	85.88
Yeast05679vs4	82.65	79.07	81.48	78.75	82.18	78.59	85.58	77.80	95.26	76.02
Vowel0	98.57	98.39	97.95	97.77	98.19	98.16	99.21	98.83	99.67	94.94
Glass016vs2	62.71	54.17	63.93	61.50	64.00	60.93	66.50	55.98	97.16	60.62
Glass2	66.54	55.30	67.50	68.28	67.43	67.76	70.16	60.38	95.71	54.24
Ecoli4	94.06	91.51	94.58	91.17	92.60	89.77	95.82	91.20	97.69	83.10
yeast1vs7	82.00	80.63	78.67	78.61	80.89	78.84	83.44	80.39	93.51	70.03
shuttle0vs4	100.00	99.12	100.00	99.57	100.00	99.57	100.00	99.57	99.99	99.97
Glass4	95.27	85.70	92.29	84.79	93.35	86.80	96.33	89.54	98.44	85.08
Page-Blocks13vs4	93.68	92.05	86.60	83.30	85.96	82.62	97.19	94.95	99.75	99.55
Abalone9-18	70.23	64.70	65.32	63.77	65.72	63.99	72.41	68.51	95.31	62.15
Glass016vs5	90.57	79.71	75.93	75.71	79.00	78.00	93.79	92.29	99.21	81.29
shuttle2vs4	95.00	90.78	100.00	98.78	98.88	95.58	98.98	96.38	99.90	99.17
Yeast1458vs7	71.25	64.65	68.25	62.76	68.63	66.10	74.80	60.78	91.58	53.67
Glass5	94.33	83.17	77.13	75.85	83.17	80.49	95.30	94.15	99.76	88.29
Yeast2vs8	78.61	77.28	77.39	77.39	77.39	77.39	79.24	77.39	91.25	80.66
Yeast4	83.58	83.15	85.42	84.34	85.65	83.53	87.15	82.18	91.01	70.04
Yeast1289vs7	74.70	77.12	76.18	75.75	76.71	76.79	80.27	78.24	94.65	68.32
Yeast5	94.68	93.58	96.48	96.49	96.70	96.63	97.46	96.01	97.77	92.33
Ecoli0137vs26	93.96	81.90	86.80	83.18	91.81	82.63	96.60	82.27	96.78	81.36
Yeast6	88.48	88.09	87.84	87.63	89.28	89.43	90.58	87.41	92.42	82.80
Abalone19	71.44	63.94	66.58	65.29	68.35	65.73	73.59	61.45	85.44	52.02
Global	85.56	81.33	83.18	81.35	84.06	81.65	88.18	**83.51**	95.46	82.17

We have selected forty-four data-sets from UCI repository [9]. The data are summarized in Table 1, showing the number of examples (#Ex.), and attributes (#Atts.), class name (minority and majority) and class attribute distribution.

In order to reduce the effect of imbalance, we will employ the SMOTE preprocessing method [13] for all our experiments balancing both classes to the 50% distribution.

We will employ the following configuration for the FRBCS: 3 labels per fuzzy partition, product T-norm as conjunction operator, together with the Penalized Certainty Factor approach for the rule weight and Fuzzy Reasoning Method of the winning rule. We have selected this fuzzy model as it achieved a good performance in previous studies for FRBCS on imbalanced data-sets [4].

The specific parameters for the genetic tuning of the amplitude are listed below:

- Number of evaluations: 5000 · number of variables.
- Population Size: 50 individuals.
- Number of Bits per Gene (for the gray codification): 30 bits.

Table 3. Wilcoxon's test to compare the IVFSs model (R^+), with and without tuning, against the Chi et al. method and C4.5 (R^-) in imbalanced data-sets

Comparison	R^+	R^-	Hypothesis ($\alpha = 0.05$)	p-value
Base IVFS				
Chi_IVFS vs. Chi_Low	506	440	Not Rejected	0.690
Chi_IVFS vs. Chi_Up	511	309	Not Rejected	0.175
Chi_IVFS vs. C4.5	435	555	Not Rejected	0.484
IVFS with Genetic Tuning of the Amplitude				
Chi_IVFS_tun vs. Chi_Low	798	148	Rejected for Chi_IVFS_tun	0.000
Chi_IVFS_tun vs. Chi_Up	596	224	Rejected for Chi_IVFS_tun	0.012
Chi_IVFS_tun vs. Chi_IVFS	642	149	Rejected for Chi_IVFS_tun	0.006
Chi_IVFS_tun vs. C4.5	611	379	Not Rejected	0.176

4.2 Analysis of the IVFSs Performance on Imbalanced Data-Sets

In the first part of our study, our aim is to analyze whether the use of the IVFSs improves the FRBCS performance by means of the comparison with the results obtained by the Chi et al.'s method, considering two amplitude values in the Data Base: using the standard MF ("lower bound" in IVFS) and a fuzzy label with a higher amplitude ("upper bound"). These results are shown in Table 2.

We observe the good behaviour of the IVFSs model, since it obtains very good results in most data-sets of the study. In order to check for significant differences between this approach and the basic FRBCSs, we carry out a Wilcoxon test (shown in Table 3) in which we observe that the rankings are very similar in all cases, concluding that the different methods have a similar performance.

When we apply the genetic tuning step, the results are enhanced considerably, obtaining the best mean result among all the algorithms of this study. The statistical analysis (also shown in Table 3) shows the goodness of this approach, since it have better behavior than the basic FRBCSs and the initial IVFSs model. Regarding C4.5, we achieve a higher ranking in this case, which implies that the IVFS with genetic tuning is a suitable methodology in order to deal with imbalanced data-sets with fuzzy models.

5 Conclusions

In this work we have analyzed the behavior of the IVFSs in the context of imbalanced data-sets. We start from an initial KB generated by a simple fuzzy rule learning method and we add a new level of fuzzy partitions in order to manage a higher knowledge for the problem.

Our experimental results have determined the goodness of this model, achieving better results than the base FRBCS. Furthermore, we have applied a post-processing step to adapt the amplitude of the upper bounds in order to contextualize this knowledge for each specific data-set by means of a genetic tuning. We have determined empirically that this methodology enhances our initial IVFSs model, outperforming the initial FRBCS and being highly competitive with the well-known C4.5 decision tree.

Acknowledgment

This work has been supported by the Spanish Ministry of Science and Technology under projects TIN2008-06681-C06-01 and TIN2007-65981.

References

1. Ishibuchi, H., Nakashima, T., Nii, M.: Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining. Springer, Heidelberg (2004)
2. Yang, Q., Wu, X.: 10 challenging problems in data mining research. International Journal of Information Technology and Decision Making 5(4), 597–604 (2006)
3. Chawla, N.V., Japkowicz, N., Kolcz, A.: Editorial: special issue on learning from imbalanced data sets. SIGKDD Explorations 6(1), 1–6 (2004)
4. Fernández, A., García, S., del Jesus, M.J., Herrera, F.: A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. Fuzzy Sets and Systems 159(18), 2378–2398 (2008)
5. Bustince, H., Montero, J., Barrenechea, E., Gomez, D.: A survey of Interval-Valued Fuzzy Sets. In: Handbook of Granular Computing. Addison-Wesley, Reading (2008)
6. Chi, Z., Yan, H., Pham, T.: Fuzzy algorithms with applications to image processing and pattern recognition. World Scientific, Singapore (1996)
7. Su, C.T., Hsiao, Y.H.: An evaluation of the robustness of MTS for imbalanced data. IEEE Transactions on Knowledge Data Engineering 19(10), 1321–1332 (2007)
8. Batista, G.E.A.P.A., Prati, R.C., Monard, M.C.: A study of the behaviour of several methods for balancing machine learning training data. SIGKDD Explorations 6(1), 20–29 (2004)
9. Asuncion, A., Newman, D.: UCI machine learning repository, University of California, Irvine, School of Information and Computer Sciences (2007), http://www.ics.uci.edu/~mlearn/MLRepository.html
10. Huang, J., Ling, C.X.: Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering 17(3), 299–310 (2005)
11. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30 (2006)
12. García, S., Herrera, F.: An Extension on Statistical Comparisons of Classifiers over Multiple Data Sets for all Pairwise Comparisons. Journal of Machine Learning Research 9, 2677–2694 (2008)
13. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligent Research 16, 321–357 (2002)
14. Ishibuchi, H., Yamamoto, T.: Rule weight specification in fuzzy rule-based classification systems. IEEE Transactions on Fuzzy Systems 13, 428–435 (2005)
15. Eshelman, L.J.: The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. In: Foundations of Genetic Algorithms, pp. 265–283. Morgan Kaufmann, San Francisco (1991)
16. García, S., Fernández, A., Luengo, J., Herrera, F.: A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability. Soft Computing 13(10), 959–977 (2009)