Interval-valued Fuzzy-Rough Feature Selection and Application for Handling Missing Values in Datasets

Shen, Qiang; Jensen, Richard

Published in:
8th Annual UK Workshop on Computational Intelligence (UKCI'08)

Publication date:
2008

Citation for published version (APA):
Shen, Q., & Jensen, R. (2008). Interval-valued Fuzzy-Rough Feature Selection and Application for Handling Missing Values in Datasets. In *8th Annual UK Workshop on Computational Intelligence (UKCI'08)* (pp. 59-64) http://hdl.handle.net/2160/646

Interval-valued Fuzzy-Rough Feature Selection and Application for Handling Missing Values in Datasets

Richard Jensen and Qiang Shen
Dept. of Computer Science
Aberystwyth University,
Ceredigion, Wales, UK
{rkj,qqs}@aber.ac.uk

Abstract

One of the many successful applications of rough set theory has been to the area of feature selection. The rough set ideology of using only the supplied data and no other information has many benefits, where most other methods require supplementary knowledge. Fuzzy-rough set theory has recently been proposed as an extension of this, in order to better handle the uncertainty present in real data. However, following this approach, there has been no investigation (theoretical or otherwise) into how to deal with missing values effectively, another problem encountered when using real world data. This paper proposes an extension of the fuzzy-rough feature selection methodology, based on interval-valued fuzzy sets, as a means to counter this problem via the representation of missing values in an intuitive way.

1 Introduction

Lately there has been great interest in developing computational intelligence methodologies which are capable of dealing with imprecision and uncertainty, and the resounding amount of research currently being done in the areas related to fuzzy and rough sets [12] is representative of this. The success of rough set theory is due in part to three aspects of the theory. Firstly, only the facts hidden in data are analysed. Secondly, no additional information about the data is required for data analysis such as thresholds or expert knowledge on a particular domain. Thirdly, it finds a minimal knowledge representation for data. As rough set theory handles only one type of imperfection found in data, it is complementary to other concepts for the purpose, such as fuzzy set theory. The two fields may be considered analogous in the sense that both can tolerate inconsistency and uncertainty - the difference being the type of uncertainty and their approach to it; fuzzy sets are concerned with vagueness, rough sets are concerned with indiscernibility.

Many deep relationships have been established and more so, most of the recent studies have concluded at this complementary nature of the two methodologies, especially in the context of granular computing [1]. Therefore, it is desirable to extend and hybridize the underlying concepts to deal with additional aspects of data imperfection. Such developments offer a high degree of flexibility and provide robust solutions and advanced tools for data analysis [9]. However, there has been no investigation into how such hybridizations may model and cope with missing values in datasets. This is a severely limiting factor for the application of these powerful techniques. In this paper, a further extension to fuzzy-rough set theory is proposed, interval-valued fuzzy-rough sets, in order to alleviate this problem. As a result of this, a new feature selection method is developed that not only handles missing values, but also alleviates the problem of defining overly-specific type-1 fuzzy similarity relations (i.e. those with crisp membership functions) through the use of interval-valued fuzzy sets.

The remainder of this paper is structured as follows. In Section 2, the fuzzy-rough hybridisation process is reviewed by briefly recalling its ingredients (fuzzy sets and rough sets) as well as its resulting end products (fuzzy-rough sets). Section 3 focuses on the proposed approach, interval-valued feature selection. Initial experimental results that demonstrate the potential of the approach are presented in Section 4. Finally, Section 5 concludes the paper and outlines some ideas for future work.

2 Theoretical Background

2.1 Fuzzy Sets

Recall that a fuzzy set [14] in $\mathbb{U}$ is an $\mathbb{U} \rightarrow [0,1]$ mapping, while a fuzzy relation in $\mathbb{U}$ is a fuzzy set in $\mathbb{U} \times \mathbb{U}$. For all y in $\mathbb{U}$, the R-foreset of y is the fuzzy set Ry defined by $\mu_{Ry}(x) = \mu_R(x,y)$ for all x in $\mathbb{U}$. If R is reflexive and symmetric, i.e., $\mu_R(x,x) = 1$ and $\mu_R(x,y) = \mu_R(y,x)$ hold for all x and y in $\mathbb{U}$, then R is called a fuzzy tolerance relation.

Fuzzy logic connectives play an important role in the hybridisation process. A triangular norm (t-norm for short) $\mathcal{T}$ is any increasing, commutative and associative $[0,1]^2 \rightarrow [0,1]$ mapping satisfying $\mathcal{T}(1,x) = x$, for all x in $[0,1]$. Common examples of t-norms include the minimum, the product and $\mathcal{T}_L$ defined by $\mathcal{T}_L(x,y) = \max(0, x+y-1)$ for x,y in [0,1]. An implicator is any $[0,1]^2 \rightarrow [0,1]$-mapping $\mathcal{I}$ that is decreasing in its first, and increasing in its second component, and that satisfies $\mathcal{I}(0,0) = 1$ and $\mathcal{I}(1,x) = x$, for all x in $[0,1]$.

2.1.1 Interval-valued Fuzzy Sets

An interval-valued fuzzy set $\widetilde{A}$ in $\mathbb{U}$ is an ordered triple of the form

$$\widetilde{A} = \{\langle x, \mu_{A_*}(x), \mu_{A^*}(x)\rangle | x \in \mathbb{U}\} \quad (1)$$

where $\mu_{A_*}(x), \mu_{A^*} \in [0,1]$ are the lower and upper membership functions which satisfy $0 \leq \mu_{A_*}(x) \leq \mu_{A^*}(x) \leq 1$, $\forall x \in \mathbb{U}$. The lower and upper membership functions correspond to the lower and upper bound of a closed interval describing the membership of x to $\widetilde{A}$.

2.2 Rough Sets

Let $I = (\mathbb{U}, \mathbb{A})$ be an information system, where $\mathbb{U}$ is a non-empty set of finite objects (the universe of discourse) and $\mathbb{A}$ is a non-empty finite set of attributes such that $a : \mathbb{U} \rightarrow V_a$ for every $a \in \mathbb{A}$. V_a is the set of values that attribute a may take. With any $P \subseteq \mathbb{A}$ there is an associated equivalence relation $IND(P)$:

$$IND(P) = \{(x,y) \in \mathbb{U}^2 | \forall a \in P,\ a(x) = a(y)\} \quad (2)$$

The partition of $\mathbb{U}$, generated by *IND(P)* is denoted $\mathbb{U}$/*IND(P)* (or $\mathbb{U}/P$ for simplicity) and can be calculated as follows:

$$\mathbb{U}/IND(P) = \otimes\{\mathbb{U}/IND(\{a\}) | a \in P\}, \quad (3)$$

where $\otimes$ is specifically defined as follows for sets A and B:

$$A \otimes B = \{X \cap Y | X \in A, Y \in B, X \cap Y \neq \emptyset\} \quad (4)$$

If $(x,y) \in IND(P)$, then x and y are indiscernible by attributes from P. The equivalence classes of the P-indiscernibility relation are denoted $[x]_P$.

Let $X \subseteq \mathbb{U}$. X can be approximated using only the information contained within P by constructing the P-*lower* and P-*upper* approximations of X:

$$\underline{P}X = \{x \in \mathbb{U} | [x]_P \subseteq X\} \quad (5)$$

$$\overline{P}X = \{x \in \mathbb{U} | [x]_P \cap X \neq \emptyset\} \quad (6)$$

The tuple $\langle \underline{P}X, \overline{P}X \rangle$ is called a rough set.

2.2.1 Feature Selection

Let P and Q be sets of attributes inducing equivalence relations over $\mathbb{U}$, then the positive region can be defined as:

$$POS_P(Q) = \bigcup_{X \in \mathbb{U}/Q} \underline{P}X \quad (7)$$

The positive region contains all objects of $\mathbb{U}$ that can be classified to classes of $\mathbb{U}/Q$ using the information in attributes P. Based on this definition, dependencies between attributes can be determined. For P, $Q \subset \mathbb{A}$, it is said that Q depends on P in a degree k ($0 \leq k \leq 1$), denoted $P \Rightarrow_k Q$, if

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|\mathbb{U}|} \quad (8)$$

If $k = 1$, Q depends totally on P, if $0 < k < 1$, Q depends partially (in a degree k) on P, and if $k = 0$ then Q does not depend on P.

The reduction of attributes is achieved by comparing equivalence relations generated by sets of attributes. Attributes are removed so that the reduced set provides the same predictive capability of the decision attribute as the original. A *reduct* R_{min} is defined as a minimal subset R of the initial attribute set $\mathbb{C}$ such that for a given set of attributes D, $\gamma_R(\mathbb{D}) = \gamma_{\mathbb{C}}(\mathbb{D})$. From the literature, R is a minimal subset if $\gamma_{R-\{a\}}(\mathbb{D}) \neq \gamma_R(\mathbb{D})$ for all $a \in R$. This means that no attributes can be removed from the subset without affecting the dependency degree. Hence, a minimal subset by this definition may not be the *global* minimum (a reduct of smallest cardinality). The goal of rough set-based feature selection is to discover reducts of smallest cardinality.

2.3 Fuzzy-Rough Sets

It is not possible in the original rough set theory to say whether two attribute values are similar and to what extent they are the same; for example, two close values may only differ as a result of noise, but in rough set theory they are considered to be as different as two values of a different order of magnitude. It is, therefore, desirable to develop techniques to provide the means of knowledge modelling for crisp and real-value attributed datasets which utilises the extent to which values are similar. This can be achieved through the use of *fuzzy-rough* sets [6]. Fuzzy-rough sets encapsulate the related but distinct concepts of vagueness (for fuzzy sets) and indiscernibility (for rough sets), both of which occur as a result of uncertainty in knowledge

Definitions for the fuzzy lower and upper approximations can be found in [13], where a T-transitive fuzzy similarity relation is used to approximate a fuzzy concept X:

$$\mu_{\underline{R_P}X}(x) = \inf_{y \in \mathbb{U}} \mathcal{I}(\mu_{R_P}(x,y), \mu_X(y)) \quad (9)$$

$$\mu_{\overline{R_P}X}(x) = \sup_{y \in \mathbb{U}} \mathcal{T}(\mu_{R_P}(x,y), \mu_X(y)) \quad (10)$$

Here, $\mathcal{I}$ is a fuzzy implicator and $\mathcal{T}$ a t-norm. R_P is the fuzzy similarity relation induced by the subset of features P:

$$\mu_{R_P}(x,y) = \mathcal{T}_{a \in P}\{\mu_{R_a}(x,y)\} \quad (11)$$

$\mu_{R_a}(x,y)$ is the degree to which objects x and y are similar for feature a, and may be defined in many ways, for example:

$$\mu_{R_a}(x,y) = 1 - \frac{|a(x) - a(y)|}{|a_{max} - a_{min}|} \quad (12)$$

$$\mu_{R_a}(x,y) = \exp(-\frac{(a(x) - a(y))^2}{2{\sigma_a}^2}) \quad (13)$$

$$\mu_{R_a}(x,y) = \max(\min(\frac{(a(y) - (a(x) - \sigma_a))}{(a(x) - (a(x) - \sigma_a))}, \frac{((a(x) + \sigma_a) - a(y))}{((a(x) + \sigma_a) - a(x))}, 0) \quad (14)$$

where ${\sigma_a}^2$ is the variance of feature a. As these relations do not necessarily display $\mathcal{T}$-transitivity, the fuzzy transitive closure can be computed for each attribute. The choice of relation is largely determined by the intended application. For feature selection, a relation such as (14) may be appropriate as this permits only small differences between attribute values of differing objects. For classification tasks, such as fuzzy-rough nearest neighbours [8], a more gradual and inclusive relation such as (12) should be used.

2.3.1 Feature Selection

In a similar way to the original crisp rough set approach, the fuzzy positive region can be defined as [10]:

$$\mu_{POS_{R_P}(\mathbb{D})}(x) = \sup_{X \in \mathbb{U}/\mathbb{D}} \mu_{\underline{R_P}X}(x) \quad (15)$$

An important issue in data analysis is discovering dependencies between attributes. The fuzzy-rough degree of dependency of $\mathbb{D}$ on the attribute subset P can be defined in the following way:

$$\gamma'_P(\mathbb{D}) = \frac{\sum_{x \in \mathbb{U}} \mu_{POS_{R_P}(\mathbb{D})}(x)}{|\mathbb{U}|} \quad (16)$$

A fuzzy-rough reduct R can be defined as a minimal subset of features that preserves the dependency degree of the entire dataset, i.e. $\gamma'_R(\mathbb{D}) = \gamma'_{\mathbb{C}}(\mathbb{D})$. Based on this, a fuzzy-rough greedy hill-climbing algorithm can be constructed that uses equation (16) to gauge subset quality. In [10], it has been shown that the dependency function is monotonic and that fuzzy discernibility matrices may also be used to discover reducts. However, there is no mechanism for modelling missing values in this framework, and is therefore limited in its application to real world datasets. This motivates the work proposed in the following section.

3 Interval-valued FRFS

Central to traditional fuzzy-rough feature selection is the fuzzy tolerance relation. From this, the fuzzy-rough lower approximations are constructed which then form the fuzzy positive regions utilised in the degree of dependency measure. Thus, the starting point for the process, type-1 fuzzy tolerance, is critical for its success. It is recognised that type-1 approaches are unable to address particular types of uncertainty due to their requirement of totally crisp membership functions [11]. An interval-valued approach may therefore be able to better handle this uncertainty and at the same time model the uncertainty inherent in missing values. Currently, there is no way to handle such values in fuzzy-rough set theory. Thus, the starting point for the proposed work is the interval-valued tolerance relation $\widetilde{R_a}(x,y)$. The constituent fuzzy relations for individual attributes can be defined via an upper (${R_a}^*$) and lower (${R_a}_*$) membership function, for example:

$$\mu_{R_{a*}}(x,y) = 1-\left(\frac{|a(x)-a(y)|}{|a_{\max}-a_{\min}|}\right)^m \quad (17)$$

$$\mu_{R_a{}^*}(x,y) = 1-\frac{|a(x)-a(y)|}{|a_{\max}-a_{\min}|} \quad (18)$$

for $m \in (0,1)$. If $m = 1$, equation (17) degenerates to a standard type-1 fuzzy tolerance relation. As with type-1 fuzzy-rough feature selection, composition of relations is achieved by conjunctively combining the individual fuzzy relations $\widetilde{R_a}$ with a t-norm $\mathcal{T}$:

$$\begin{aligned}\mu_{\widetilde{R_P}}(x,y) &= \mathcal{T}_{a\in P}\{\mu_{\widetilde{R_a}}(x,y)\} \quad (19)\\ &= [\mathcal{T}_{a\in P}\{\mu_{R_{a*}}(x,y)\}, \mathcal{T}_{a\in P}\{\mu_{R_a{}^*}(x,y)\}]\end{aligned}$$

Based on the definitions above, the interval-valued P-lower and P-upper approximation of a concept $\widetilde{X}$ are here defined as

$$\mu_{\widetilde{\underline{R_P}X}}(x) = \inf_{y\in\mathbb{U}} \mathcal{I}(\mu_{\widetilde{R_P}}(x,y), \mu_{\widetilde{X}}(y)) \quad (20)$$

$$\mu_{\widetilde{\overline{R_P}X}}(x) = \sup_{y\in\mathbb{U}} \mathcal{T}(\mu_{\widetilde{R_P}}(x,y), \mu_{\widetilde{X}}(y)) \quad (21)$$

where $\widetilde{R_P}(x,y)$ is an interval-valued fuzzy tolerance relation. The tuple $\langle\widetilde{\underline{R_P}X}, \widetilde{\overline{R_P}X}\rangle$ is called an interval-valued fuzzy-rough set.

3.1 Feature Selection

For the purposes of feature selection, the work presented in this paper only considers the use of the interval-valued lower approximation:

$$\begin{aligned}\mu_{\widetilde{\underline{R_P}X}}(x) &= \inf_{y\in\mathbb{U}} \mathcal{I}(\mu_{\widetilde{R_P}}(x,y), \mu_{\widetilde{X}}(x))\\ &= \inf_{y\in\mathbb{U}} [\mathcal{I}(\mu_{R_P{}^*}(x,y), \mu_{X_*}(x)),\\ &\qquad \mathcal{I}(\mu_{R_{P*}}(x,y), \mu_{X^*}(x))]\end{aligned}$$

This provides a measure of the uncertainty of membership of an object to a given concept $\widetilde{X}$ as a result of the underlying uncertainty in the similarity between this object and others in the universe. Note that the use of $\mu_{R_P{}^*}$ and $\mu_{R_{P*}}$ is reversed due to the properties of fuzzy implication. The resulting interval is a lower and upper bound on the true membership degree. Based on this, the interval-valued positive region is defined:

$$\mu_{\widetilde{POS_P(\mathbb{D})}}(x) = \sup_{\widetilde{X}\in\mathbb{U}/\mathbb{D}} \mu_{\widetilde{\underline{R_P}X}}(x) \quad (22)$$

From this, the interval-valued degree of dependency of decision features $\mathbb{D}$ on a feature subset P is defined as:

$$\tilde{\gamma}_P(\mathbb{D}) = \frac{|\widetilde{POS_P(\mathbb{D})}|}{|\mathbb{U}|} \quad (23)$$

$$= \left[\sum_{y\in\mathbb{U}} \frac{POS_{P*}(\mathbb{D})(y)}{|\mathbb{U}|}, \sum_{y\in\mathbb{U}} \frac{POS_P{}^*(\mathbb{D})(y)}{|\mathbb{U}|}\right]$$

In [5], a normalised version of dependency was introduced for FRFS. The equivalent interval-valued normalised version is as follows:

$$\tilde{\gamma}_P(\mathbb{D}) = \frac{|\widetilde{POS_P(\mathbb{D})}|}{|\widetilde{POS_\mathbb{C}(\mathbb{D})}|} \quad (24)$$

$$= \left[\frac{1}{\mathbb{U}}\sum_{y\in\mathbb{U}} \frac{POS_{P*}(\mathbb{D})(y)}{POS_{\mathbb{C}*}(\mathbb{D})(y)}, \frac{1}{\mathbb{U}}\sum_{y\in\mathbb{U}} \frac{POS_P{}^*(\mathbb{D})(y)}{POS_\mathbb{C}{}^*(\mathbb{D})(y)}\right]$$

Here, $\tilde{\gamma}_P(\mathbb{D})$ is an interval $[\gamma_{P*}(\mathbb{D}), \gamma_P{}^*(\mathbb{D})]$ that describes the extent to which the features in P are predictive of the decision feature(s). P is called a fuzzy decision superreduct to degree α if $\tilde{\gamma}_P(\mathbb{D}) \geq \alpha$, and a fuzzy decision reduct to degree α if moreover for all $P' \subset P$, $\tilde{\gamma}_{P'}(\mathbb{D}) < \alpha$. Note that if a type-1 fuzzy tolerance relation is used, then these definitions degenerate to their traditional fuzzy-rough counterparts. Core features (i.e. those that cannot be removed without introducing inconsistencies) may be determined by considering the change in dependency of the full set of conditional features when individual attributes are removed:

$$Core(\mathbb{C}) = \{a \in \mathbb{C} | \tilde{\gamma}_{\mathbb{C}-\{a\}}(\mathbb{D}) \neq \tilde{\gamma}_\mathbb{C}(\mathbb{D})\} \quad (25)$$

Subset search can be conducted by whichever mechanism is appropriate; for example, greedy hill-climbing, genetic algorithms [2], ant colony optimization [7] etc. Here, the standard greedy hill-climbing approach is adopted. The dependency degree is monotonic in both the lower and upper bounds, and so search continues until $[1,1]$ is reached (no uncertainty) or there is no improvement in dependency.

3.2 Missing Values

The underlying interval-valued tolerance relation can be modified in order to model the uncertainty resulting from missing values. If an object contains a missing value for a particular feature, then the resulting degree of similarity with other objects is unknown. In an interval-valued context, this can be modelled by returning the unit interval when an attribute value is missing for one or both objects:

$$\mu_{\widetilde{R_a}}(x,y) = \begin{cases} \mu_{\widetilde{R_a}}(x,y) & \text{if } x, y \neq *, \\ [0,1] & \text{otherwise} \end{cases} \tag{26}$$

where missing values are denoted by $*$. Again, relations are composed via a t-norm. The resulting interval-valued lower approximations, positive regions and dependency can then be used to gauge subset quality in an identical manner to that of the approach in section 3.1. The resulting algorithm is called FRFS-II.

4 Initial experimentation

To test the robustness of the proposed approach in the presence of missing values four benchmark datasets were used, obtained from [3] and containing no missing values initially. Values were randomly corrupted to become missing based on a supplied probability of mutation, and the FRFS-II algorithm run, using $m = 0.9$ for the tolerance relation. This was carried out ten times for each dataset. Table 1 shows the dataset details as well as the probabilities of value corruption and the resulting ranges of numbers of missing values; e.g. for the water 2 dataset and probability of mutation 0.0005%, datasets were produced with 4 to 14 missing values present. For comparison, the type-1 FRFS algorithm was run on the original, uncorrupted data. As the corruption probability increases, the amount of missing data greatly increases, with the final column showing a degree of missing values not often seen in real data but useful to test the robustness of the approach.

The resulting reduct sizes (averaged over repeated runs) can be seen in table 2. All discovered reducts produced a dependency degree of 1 for the uncorrupted data and hence were all fuzzy-rough reducts. This shows the resilience of the approach when faced with corrupted data, but may not necessarily be the case in general as it will depend on the extent of missing values and similarity relation chosen. For small numbers of missing values, FRFS-II manages to locate identical or equivalent reducts to the original FRFS approach (which was applied to the uncorrupted data). As the extent of data corruption increases, the task becomes more difficult, with FRFS-II selecting more features to compensate for the lack of information but still producing valid fuzzy-rough reducts.

5 Conclusion

This paper proposed an interval-valued approach to fuzzy-rough feature selection that successfully handles the uncertainty that cannot be modelled by a type-1 approach. In particular, this method can handle missing values effectively and in an intuitive way.

Future work will involve further experimental investigations, particularly to see if the trend observed in this paper is maintained for other datasets. This will include an analysis of the impact of the choice of similarity relation and parameter m. There will also need to be an evaluation of the effectiveness of the discovered reducts for the task of classification. Recent work in the area of type-1 FRFS for this purpose has shown that such reducts improve performance [10] and a similar improvement is expected here. Additionally, the work in [4] proposed novel t-norms and implicators for interval-valued fuzzy sets; these should be of great benefit for interval-valued FRFS.

References

[1] A. Bargiela, W. Pedrycz, *Granular Computing. An introduction.* Kluwer Academic Publishers, 2002.

[2] A.T. Bjorvand and J. Komorowski, Practical Applications of Genetic Algorithms for Efficient Reduct Computation, *Proc. 15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics*, A. Sydow, ed., vol. 4, pp. 601–606, 1997.

[3] C.L. Blake, C.J. Merz, *UCI Repository of Machine Learning Databases.* Irvine, University of California, 1998. http://archive.ics.uci.edu/ml/

[4] C. Cornelis, G. Deschrijver, E.E. Kerre, Advances and challenges in interval-valued fuzzy logic, *Fuzzy Sets and Systems*, vol. 157, no. 5, pp. 622–627, 2006.

[5] C. Cornelis, G. Hurtado Martín, R. Jensen, D. Ślęzak, Feature Selection with Fuzzy Decision Reducts, *3rd Int. Conf. on Rough Sets and Knowledge Technology (RSKT'08)*, pp. 284–291, 2008.

[6] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, *International Journal of General Systems*, vol. 17, 91–209, 1990.

Table 1: Number of missing values

Dataset	Objects	Features	Missing values (range)				
			0.0005%	0.001%	0.005%	0.01%	0.1%
Heart	270	14	1-3	1-6	12-24	26-43	321-387
Water 2	390	39	4-14	9-30	62-88	123-175	1446-1533
Water 3	390	39	3-13	12-23	65-98	128-163	1462-1531
Wine	178	14	1-5	1-4	7-15	20-30	206-246

Table 2: Reduct size for various data corruption probabilities

Dataset	Features	Original FRFS	Average reduct size				
			0.0005%	0.001%	0.005%	0.01%	0.1%
Heart	14	8	8	8	9.1	10.3	14
Water 2	39	7	7	7	7	7.2	11.5
Water 3	39	7	7	7	7.2	7.9	11.8
Wine	14	6	6	6	6	6	9.1

[7] R. Jensen, Performing Feature Selection with ACO, *Swarm Intelligence and Data Mining*, A. Abraham, C. Grosan and V. Ramos (eds.), Studies in Computational Intelligence, vol. 34, pp. 45–73, 2006.

[8] R. Jensen, C. Cornelis, A New Approach to Fuzzy-Rough Nearest Neighbour Classification, Proceedings of the *6th Int. Conf. on Rough Sets and Current Trends in Computing*, to appear, 2008.

[9] R. Jensen, Q. Shen, *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*, IEEE Press/Wiley & Sons, 2008.

[10] R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selection, *IEEE Transactions on Fuzzy Systems*, to appear, 2008.

[11] J.M. Mendel, R.I. John, Type-2 Fuzzy Sets Made Simple, *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, 117-127, 2002.

[12] Z. Pawlak, Rough sets, *International Journal of Computer and Information Sciences*, vol. 11(5), 341–356, 1982.

[13] A.M. Radzikowska, E.E. Kerre, E.E., A comparative study of fuzzy rough sets, *Fuzzy Sets and Systems*, vol. 126, 137–156, 2002.

[14] L.A. Zadeh, Fuzzy sets, *Information and Control*, vol. 8, 338–353, 1965.