Fuzzy-Rough Sets for Descriptive Dimensionality Reduction

Richard Jensen and Qiang Shen
{richjens,qiangs}@dai.ed.ac.uk
Centre for Intelligent Systems and their Applications
Division of Informatics, The University of Edinburgh

Abstract - One of the main obstacles facing current fuzzy modelling techniques is that of dataset dimensionality. To enable these techniques to be effective, a redundancy-removing step is usually carried out beforehand. Rough Set Theory (RST) has been used as such a dataset pre-processor with much success, however it is reliant upon a *crisp* dataset; important information may be lost as a result of quantization. This paper proposes a new dimensionality reduction technique that employs a hybrid variant of rough sets, *fuzzy-rough* sets, to avoid this information loss.

I. Introduction

It is estimated that every 20 months or so the amount of information in the world doubles. In the same way, tools for use in the various knowledge fields (acquisition, storage, retrieval, maintenance, etc) must develop to combat this growth. Knowledge is only valuable when it can be used efficiently and effectively [1].

Over the past twenty years, rough set theory (RST) has become a topic of great interest to researchers and has been applied to many domains. This success is due in part to the following aspects of the theory: only the facts hidden in data are analysed, no additional information about the data is required, and it helps find a minimal knowledge representation. One particular use of RST is that of attribute reduction in datasets. Given a dataset with discretized attribute values, it is possible to find a subset of the original attributes that are the most informative (termed a *reduct*); all other attributes can be removed from the dataset with minimal information loss. This method tends to be a pre-processing step to reduce dataset dimensionality before some other action is performed (for example, induction of rules [9]).

However, it is most often the case that the values of attributes may be both symbolic and *real-valued*, and this is where traditional rough set theory encounters a problem. It is not possible in the 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 RST they are considered to be as different as two values of a different order of magnitude.

One answer to this problem has been to discretize the dataset beforehand, producing a new dataset with crisp values. This is often still inadequate, however, as the degrees of membership of values to discretized values are not considered at all. For example, two values may both be mapped to the same class "Negative", but one may be much more negative than the other when comparing the two underlying real values. This is a source of information loss, which is against the rough set ideology of retaining information content. It is clear that there is a need for some method that will provide the means of data reduction for crisp and real-value attributed datasets which utilises the extent to which values are similar. This could be achieved through the use of *fuzzy-rough* sets [2]. 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 or data; a method employing fuzzy-rough sets should be better equipped to handle this uncertainty.

The rest of this paper is structured as follows. Section 2 discusses the fundamentals of rough set theory, in particular focusing on dimensionality reduction. The third section introduces the hybrids of rough and fuzzy sets; section 4 builds on these definitions to outline a procedure for fuzzy-rough dimensionality reduction. This is then applied to a test domain and compared with traditional crisp RSAR, demonstrating the power of the new approach. Finally, the paper concludes with a discussion of the results and outlines future work to be carried out.

II. Theoretical Foundations

A. Rough Sets

A rough set [7] is an approximation of a vague concept by a pair of precise concepts, called lower and upper approximations (which are informally a classification of the domain of interest into disjoint categories). Objects be-

longing to the same category characterised by the same attributes (or features) are not distinguishable.

The Rough Set Attribute Reduction (RSAR) method [9] employs rough sets to remove redundant conditional attributes from discrete-valued datasets, while retaining their information content. Central to RSAR is the concept of indiscernibility. Let $I = (U, A)$ be an information system, where U is a non-empty set of finite objects (the universe); A is a non-empty finite set of attributes such that $a : U \rightarrow V_a$ for every $a \in A$; V_a is the value set for attribute a. In a decision system, $A = \{C \cup D\}$ where C is the set of conditional attributes and D is the set of decision attributes.

TABLE I
EXAMPLE DATASET

Object	A	B	C	Q
1	-0.4	-0.3	-0.5	no
2	-0.4	0.2	-0.1	yes
3	-0.3	-0.4	-0.3	no
4	0.3	-0.3	0	yes
5	0.2	-0.3	0	yes
6	0.2	0	0	no

With any $P \subseteq A$ there is an associated equivalence relation $IND(P)$:

$$IND(P) = \{(x, y) \in U^2 \mid \forall\, a \in P\ a(x) = a(y)\} \quad (1)$$

If $(x, y) \in IND(P)$, then x and y are indiscernible by attributes from P. The partition of U, generated by $IND(P)$ is denoted U/P and can be calculated as follows:

$$U/P = \otimes\{a \in P : U/IND(\{a\})\}, \quad (2)$$

where

$$A \otimes B = \{X \cap Y : \forall X \in A, \forall Y \in B, X \cap Y \neq \emptyset\} \quad (3)$$

TABLE II
DISCRETIZED DATASET

Object	A	B	C	Q
1	N_A	N_B	N_C	no
2	N_A	Z_B	Z_C	yes
3	N_A	N_B	N_C	no
4	Z_A	N_B	Z_C	yes
5	Z_A	N_B	Z_C	yes
6	Z_A	Z_B	Z_C	no

To illustrate the operation of RSAR, an example dataset (table I) has been quantized, producing the discretized dataset seen in table II. Partitioning the universe of objects according to attributes C and Q gives

$U/C = \{\{1,3\}\{2,4,5,6\}\}, \quad U/Q = \{\{1,3,6\}\{2,4,5\}\}$

The equivalence classes of the P-indiscernibility relation are denoted $[x]_P$. Let $X \subseteq U$, the P-lower approximation of a set can now be defined as:

$$\underline{P}X = \{x \mid [x]_P \subseteq X\} \quad (4)$$

Let P and Q be equivalence relations over U, then the positive region can be defined as:

$$POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}X \quad (5)$$

The positive region contains all objects of U that can be classified to classes of U/Q using the knowledge in attributes P. In the example, the C-lower approximation is calculated for each equivalence class of U/Q to give the C-positive region.

$\underline{C}X = \{x \mid [x]_C \subseteq X\}$
$\underline{C}\{1, 3, 6\} = \{1, 3\}, \quad \underline{C}\{2, 4, 5\} = \emptyset$
$POS_C(Q) = \bigcup_{X \in U/Q} \underline{C}X = \{1, 3\}$

An important issue in data analysis is discovering dependencies between attributes. Intuitively, a set of attributes Q depends totally on a set of attributes P, denoted $P \Rightarrow Q$, if all attribute values from Q can be uniquely determined by values of attributes from P. In particular, if there exists a functional dependency between values of Q and P, then Q depends totally on P. Dependency can be defined in the following way:

For $P, Q \subset A$, Q depends on P in a degree k ($0 \leq k \leq 1$), denoted $P \Rightarrow_k Q$, if

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|} \quad (6)$$

If $k = 1$ Q depends totally on P, if $k < 1$ Q depends partially (in a degree k) on P, and if $k = 0$ Q does not depend on P. The dependency for the example is

$\gamma_C(Q) = \frac{|\{1,3\}|}{|U|} = 2/6$

By calculating the change in dependency when an attribute is removed from the set of considered conditional attributes, a measure of the significance of the attribute can be obtained. The higher the change in dependency, the more significant the attribute is.

B. Reducts

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 quality of classification as the original. A *reduct* is defined as a subset R of the conditional attribute set C

such that $\gamma_R(D) = \gamma_C(D)$. A given dataset may have many attribute reduct sets, so the set R of all reducts is defined as:

$$R = \{X : X \subseteq C, \gamma_X(D) = \gamma_C(D)\} \quad (7)$$

The intersection of all the sets in R is called the *core*, the elements of which are those attributes that cannot be eliminated without introducing more contradictions to the dataset. In RSAR, a reduct with minimum cardinality is searched for; in other words an attempt is made to locate a single element of the minimal reduct set $R_{min} \subseteq R$:

$$R_{min} = \{X : X \in R, \forall Y \in R, |X| \le |Y|\} \quad (8)$$

A basic way of achieving this is to calculate the dependencies of all possible subsets of C. Any subset X with $\gamma_X(D) = 1$ is a reduct; the smallest subset with this property is a minimal reduct. However, for large datasets this method is impractical and an alternative strategy is required.

1. $R \leftarrow \{\}$
2. do
3. $T \leftarrow R$
4. $\forall x \in (C - R)$
5.   if $\gamma_{R \cup \{x\}}(D) > \gamma_T(D)$
6. $T \leftarrow R \cup \{x\}$
7. $R \leftarrow T$
8. until $\gamma_R(D) = \gamma_C(D)$
9. return R

Fig. 1. The QUICKREDUCT Algorithm

The QUICKREDUCT algorithm [4], [9] attempts to calculate a minimal reduct without exhaustively generating all possible subsets. It starts off with an empty set and adds in turn those attributes that result in the greatest increase in $\gamma_P(Q)$, until this produces its maximum possible value for the dataset (usually 1).

Running the algorithm on the example dataset results in a non-minimal reduct $\{A, B, C\}$ (i.e. *all* attributes are chosen by this method). It can be seen upon examination of the dataset that the set of attributes $\{A, B\}$ is the true minimal reduct. Although γ is a non-optimal heuristic, the algorithm has also suffered due to the information lost in the discretization procedure. A method using fuzzy-rough sets should therefore be more informed.

III. Fuzzy-Rough Hybrids

A. Fuzzy Equivalence Classes

In the same way that crisp equivalence classes are central to rough sets, *fuzzy* equivalence classes are central to the

hybrid approaches. Rough set theory can be extended by the inclusion of a similarity relation [2] on the universe, a fuzzy set S which determines the extent to which two elements are similar in S. The usual properties of reflexivity ($\mu_S(x, x) = 1$), symmetry ($\mu_S(x, y) = \mu_S(y, x)$) and transitivity ($\mu_S(x, z) \ge \mu_S(x, y) \wedge \mu_S(y, z)$) hold.

Using the fuzzy similarity relation, it is now possible to define the fuzzy equivalence class $[x]_S$ for objects close to x:

$$\mu_{[x]_S}(y) = \mu_S(x, y) \quad \forall y \in X \quad (9)$$

This definition degenerates to the normal definition of equivalence classes when S is non-fuzzy. The following axioms should hold for a fuzzy equivalence class F_i [3]:

1. μ_{F_i} is normalized, $\exists x, \mu_{F_i}(x) = 1$

2. $\mu_{F_i}(x) \wedge \mu_S(x, y) \le \mu_{F_i}(y)$

3. $\mu_{F_i}(x) \wedge \mu_{F_i}(y) \le \mu_S(x, y)$

The first axiom corresponds to the requirement that an equivalence class is non-empty. The second axiom states that elements in y's neighbourhood are in the equivalence class of y. The final axiom states that any two elements in F_i are related via S. With these definitions, the concepts of rough-fuzzy and fuzzy-rough sets may be introduced.

B. Rough-Fuzzy Sets

A rough-fuzzy set is a generalisation of a rough set, derived from the approximation of a fuzzy set in a crisp approximation space. In RSAR, this corresponds to the case when the decision attribute(s) values are fuzzy. The lower and upper approximations incorporate the extent to which objects belong to these sets, and are defined as:

$$\mu_{\underline{R}X}([x]_R) = inf\{\mu_X(x)|x \in [x]_R\} \quad (10)$$

$$\mu_{\overline{R}X}([x]_R) = sup\{\mu_X(x)|x \in [x]_R\} \quad (11)$$

where $\mu_X(x)$ is the degree to which x belongs to fuzzy equivalence class X, and each $[x]_R$ is crisp. The tuple $< \underline{R}X, \overline{R}X >$ is called a rough-fuzzy set. It can be seen that in the crisp case (where $\mu_X(x)$ is 1 or 0), the above definitions become identical to that of the traditional lower and upper approximations.

C. Fuzzy-Rough Sets

Rough-fuzzy sets can be generalised to fuzzy-rough sets [2], where *all* equivalence classes may be fuzzy. In RSAR, this means that the decision values and the conditional

values may all be fuzzy. The lower and upper approximations are now:

$$\mu_{\underline{X}}(F_i) = inf_x max\{1 - \mu_{F_i}(x), \mu_X(x)\} \quad \forall i \qquad (12)$$

$$\mu_{\overline{X}}(F_i) = sup_x min\{\mu_{F_i}(x), \mu_X(x)\} \quad \forall i \qquad (13)$$

where F_i denotes a single fuzzy equivalence class. The tuple $< \underline{X}, \overline{X} >$ is called a fuzzy-rough set. Again, it can be seen that these definitions degenerate to traditional rough sets when all equivalence classes are crisp. Additionally, if all F_is are crisp, the result is a rough-fuzzy set. Fuzzy-rough sets have been used before [8] but not in the field of dimensionality reduction.

IV. Fuzzy-Rough Dimensionality Reduction

A. The Approach

Unlike its crisp counterpart, the present work on fuzzy RSAR employs fuzzy equivalence classes. Consider the crisp partitioning $U/Q = \{\{1,3,6\}\{2,4,5\}\}$ of the example. This contains two equivalence classes ($\{1,3,6\}$ and $\{2,4,5\}$) that can be thought of as fuzzy sets, with those elements belonging to the class possessing a membership of one, zero otherwise. For the first class, the objects 2, 4 and 5 have a membership of zero. Extending this to the fuzzy case is straightforward: objects can be allowed to assume membership values in the interval [0,1]. As Q contains discrete values, its equivalence classes can be determined in the traditional way. However, U/Q is not restricted to crisp partitions only; fuzzy partitions are equally acceptable.

As before, the next step is to calculate the lower approximations of the sets. In fuzzy RSAR the approximations are themselves fuzzy and are defined as:

$$\mu_{\underline{X}}(F_i) = inf_x max\{1 - \mu_{F_i}(x), \mu_X(x)\} \; \forall i$$

This is similar to the traditional approach where the positive region is determined by the use of the equivalence classes in U/Q. The membership of each object x in the dataset to the fuzzy equivalence class F_i is obtained from the definition of F_i. Within the example, for the first equivalence class $X = \{1,3,6\}$, both $\mu_{\{1,3,6\}}(N_A)$ and $\mu_{\{1,3,6\}}(Z_A)$ need to be calculated as both N_A and Z_A are the fuzzy equivalence classes in U/A. For simplicity, consider only the first of these fuzzy sets:

$$\mu_{\{1,3,6\}}(N_A) = inf_x max\{1 - \mu_{N_A}(x), \mu_{\{1,3,6\}}(x)\}$$

The membership of each object x in the dataset to the fuzzy equivalence class N_A is obtained from the definition of N_A (figure 2). As U/Q in this case is crisp, the extent to which x belongs to $\{1,3,6\}$ is 1 only if $x \in \{1,3,6\}$, and 0 otherwise. For example, regarding object 2:
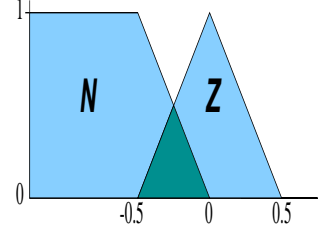

Fig. 2. Corresponding fuzzy sets

$$max\{1 - \mu_{N_A}(2), \mu_{\{1,3\}}(2)\} = max\{1 - 0.8, 0\} = 0.2$$

Calculating this for all objects in the dataset and for both fuzzy equivalence classes gives
$$\mu_{\{1,3,6\}}(N_A) = inf\{1, 0.2, 1, 1, 1, 1\} = 0.2$$
$$\mu_{\{1,3,6\}}(Z_A) = inf\{1, 0.8, 1, 0.6, 0.4, 1\} = 0.4$$

Similarly for $X = \{2,4,5\}$,
$$\mu_{\{2,4,5\}}(N_A) = inf\{0.2, 1, 0.4, 1, 1, 1\} = 0.2$$
$$\mu_{\{2,4,5\}}(Z_A) = inf\{0.8, 1, 0.6, 1, 1, 0.4\} = 0.4$$
The crisp positive region in traditional rough set theory is defined as the union of the lower approximations. Therefore, by the extension principle, the fuzzy positive region of a fuzzy equivalence class $F_i \in U/A$ can be defined as:

$$\mu_{POS_A}(F_i) = \sup_{X \in U/Q} \mu_{\underline{X}}(F_i) \qquad (14)$$

In the example, $\mu_{POS_A}(N_A) = 0.2$, and $\mu_{POS_A}(Z_A) = 0.4$. Again using the extension principle, it is possible to define the membership of an object $x \in U$ to the fuzzy positive region:

$$\mu_{POS_A}(x) = \sup_{F_i \in U/A} min(\mu_{F_i}(x), \mu_{POS_A}(F_i)) \qquad (15)$$

This states that the extent to which object x belongs to POS_A is the degree to which it belongs to each fuzzy equivalence class and the degree to which the equivalence class belongs to POS_A. So, for object 2:

$$\mu_{POS_A}(2) = max(min(0.8, 0.2), min(0.2, 0.4)) = 0.2$$

The memberships for the remaining objects are:

$\mu_{POS_A}(1) = 0.2$, $\mu_{POS_A}(3) = 0.4$, $\mu_{POS_A}(4) = 0.4$, $\mu_{POS_A}(5) = 0.4$, $\mu_{POS_A}(6) = 0.4$

Finally, using the definition of the fuzzy positive region, the new dependency function can be defined as follows

$$\gamma'_A(Q) = \frac{|\mu_{POS_A}(x)|}{|U|} = \frac{\sum_{x \in U} \mu_{POS_A}(x)}{|U|} \qquad (16)$$

As with traditional rough sets, the dependency is the proportion of objects that are discernible out of

the entire dataset using only the information in A. In the present approach, this corresponds to determining the cardinality of $\mu_{POS_A}(x)$ divided by the total number of objects in the universe. In the example this is:

$$\gamma'_A(Q) = \frac{\sum_{x \in U} \mu_{POS_A}(x)}{|U|} = 2.0/6$$

Calculating for B and C gives:

$$\gamma'_B(Q) = \frac{2.4}{6}, \quad \gamma'_C(Q) = \frac{1.6}{6}$$

If this process is to be useful, it must be able to deal with multiple attributes, finding the dependency for various subsets of the original attribute set. For example, it may be necessary to be able to determine the degree of dependency for $P = \{A, B\}$. In the crisp case, U/P contains sets of objects grouped together that are indiscernible according to both attributes A and B. In the fuzzy case, objects may belong to many equivalence classes, so the cartesian product of U/A and U/B must be considered.

$$U/P = \otimes\{a \in P : U/IND(a)\}, \qquad (17)$$

In the example case where $P = \{A, B\}$ and $U/A = \{N_A, Z_A\}$, $U/B = \{N_B, Z_B\}$:
$U/P = U/A \otimes U/B = \{N_A, Z_A\} \otimes \{N_B, Z_B\}$
$= \{\{N_A, N_B\}, \{N_A, Z_B\}, \{Z_A, N_B\}, \{Z_A, Z_B\}\}$

Each set in U/P is an equivalence class. The extent to which an object belongs to such a class is calculated by finding the conjunction of membership degrees to all sets in the class:

$$\mu_{\{F_1 \dots F_n\}}(x) = min(\mu_{F_1}(x), \mu_{F_2}(x), ..., \mu_{F_n}(x)) \quad (18)$$

Considering the first equivalence class $\{N_A, N_B\}$ in the example, the membership of an object x can be calculated by:

$$\mu_{\{N_A, N_B\}}(x) = min(\mu_{N_A}(x), \mu_{N_B}(x))$$

B. Reduct Computation

A problem can be seen here when this approach is compared to the traditional approach. In RST, a reduct is defined as a subset R of the attributes which have the same information content as the full attribute set A. In terms of the dependency function this means that the values $\gamma(R)$ and $\gamma(A)$ are identical and equal to 1 if the dataset is consistent. However, in the fuzzy-rough approach this is not necessarily the case as the uncertainty encountered when objects belong to many fuzzy equivalence classes results in a reduced total dependency.

One solution could be to stop the QUICKREDUCT algorithm if there is no gain in dependency with any of the

attributes added to the potential reduct (i.e. a "dead end" has been reached by the search). This may well fail in larger datasets as it may be the case that the algorithm will not stop until all attributes are added; each addition may produce a small increase in the overall dependency. The introduction of some threshold value to terminate the algorithm if the change in dependency is not significant enough could alleviate this problem. However, it is not known at the present whether this is a suitable solution as it requires information outside the given dataset.

With these issues in mind, a new QUICKREDUCT al-

1. $R \leftarrow \{\}$, $\gamma'_{best} \leftarrow 0$, $\gamma'_{prev} \leftarrow 0$
2. do
3. $T \leftarrow R$
4. $\gamma'_{prev} \leftarrow \gamma'_{best}$
5. $\forall x \in (C - R)$
6.   if $\gamma'_{R \cup \{x\}}(D) > \gamma'_T(D)$
7. $T \leftarrow R \cup \{x\}$
8. $\gamma'_{best} \leftarrow \gamma'_T(D)$
9. $R \leftarrow T$
10. until $\gamma'_{best} = \gamma'_{prev}$
11. return R

Fig. 3. The New QUICKREDUCT Algorithm

gorithm has been developed (figure 3). It employs the new dependency function γ' to choose which attributes to add to the current reduct candidate in the same way as the original QUICKREDUCT process. The algorithm terminates when the addition of any remaining attribute does not increase the dependency (such a criterion could be used with the original QUICKREDUCT algorithm).

To obtain a reduct from the example dataset, the algorithm first evaluates the dependency values for each attribute. For the example,

$$\gamma'_{\{A\}}(Q) = \frac{2}{6}, \quad \gamma'_{\{B\}}(Q) = \frac{2.4}{6}, \quad \gamma'_{\{C\}}(Q) = \frac{1.6}{6}$$

From this it can be seen that attribute B will cause the greatest increase in dependency degree. This attribute is chosen and added to the potential reduct.

$$\gamma'_{\{A,B\}}(Q) = \frac{3.4}{6}, \quad \gamma'_{\{B,C\}}(Q) = \frac{3.2}{6}$$

Adding attribute A to the reduct candidate causes the larger increase of dependency, so the new candidate becomes $\{A, B\}$. Lastly, attribute C is added to the reduct:

$$\gamma'_{\{A,B,C\}}(Q) = \frac{3.4}{6}$$

As this causes no increase in dependency, the algorithm stops and outputs the reduct $\{A, B\}$. Unlike crisp RSAR,

the true minimal reduct was found using the information on degrees of membership. It is clear from this example alone that the information lost by using crisp RSAR can be important when trying to discover the smallest reduct from a dataset.

V. A Test Application

A. The Problem

In order to evaluate the utility of the new fuzzy RSAR approach, a challenging test dataset was chosen, namely the Water Treatment Plant Database [6]. The dataset itself is a set of historical data charted over 521 days, with 38 different input attributes measured daily. Each day is classified into one of thirteen categories depending on the operational status of the plant. However, these can be collapsed into just two or three categories (i.e. *Normal* and *Faulty*, or *OK*, *Good* and *Faulty*) as many classifications reflect similar performance.

It is likely that not all of the 38 input attributes are required to determine the status of the plant, hence the dimensionality reduction step. However, choosing the most informative attributes is a difficult task as there will be many dependencies between subsets of attributes. There is also a monetary cost involved in monitoring these inputs, so it is desirable to reduce this number. This problem has been used recently [9] to illustrate the effectiveness of applying crisp RSAR as a pre-processing step to rule induction. However, as the dataset is entirely composed of real-valued attributes, discretization had to be performed. This clearly is a potential source of information loss. By applying the present work, such loss can be reduced.

The original monitoring system consisted of several modules; it is this modular structure that allows the new fuzzy RSAR technique to replace the existing crisp method. Originally, a precategorisation step preceded dimensionality reduction where attribute values were quantized. This is now replaced by a fuzzification procedure which leaves the underlying attribute values unchanged but generates a series of fuzzy sets for each attribute. These sets are generated entirely from the data while exploiting the statistical data attached to the dataset (in keeping with the rough set ideology). Based on these fuzzy sets and the original real-valued dataset, fuzzy RSAR calculates a reduct and reduces the dataset accordingly. Finally, fuzzy rule induction is performed on the reduced dataset using the modelling algorithm given in [5]. Note that this algorithm is not optimal, nor is the fuzzification. Yet the comparisons given below are fair due to their common background. Alternative fuzzy modelling techniques can be employed for this if available. The algorithm used herein works based on the use
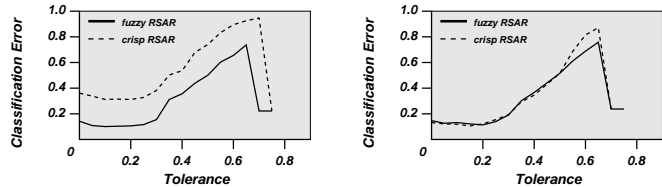

Fig. 4. Classification errors for 2- and 3-class datasets

of a tolerance threshold and produces a set of rules that enable classification. It is difficult to determine the best tolerance beforehand, so the experiments presented here are performed over a tolerance range.

B. Results

TABLE III

EXTENT OF DIMENSIONALITY REDUCTION

Method	Attributes 2-class	Attributes 3-class
Fuzzy RSAR	10	11
Crisp RSAR	11	11

Two factors must be considered when comparing fuzzy and crisp RSAR: the extent of dimensionality reduction and the resulting classification accuracy. As can be seen from table III, fuzzy RSAR selects fewer attributes than the crisp method for the 2-class dataset and results in a lower classification error over the entire tolerance range (figure 4). Both results show that there is a lot of redundancy in the dataset which may be removed with little loss in classification accuracy. For the 3-class dataset, the approaches perform similarly using the same number of attributes (but not *identical* attributes). The classification results can also be seen in figure 4.

Both approaches are dependent on an accurate fuzzification of the test dataset. Fuzzy RSAR in particular is more sensitive to this as it depends on the generated fuzzy sets to derive reductions. As indicated previously, no attempt has been made here to optimise the algorithm and fuzzification; however, it can be expected that this would help improve the results.

VI. Conclusion

This paper has highlighted the shortcomings of traditional rough set attribute reduction when applied to datasets with real-valued attributes and has proposed a new method based on fuzzy-rough sets. The new approach incorporates the information usually lost in crisp discretization by utilising the generated fuzzy-rough sets to provide a more informed technique. This has been

illustrated with a test application. The results show that fuzzy RSAR reduces dataset dimensionality without much loss in information content. The approach performs better than traditional RSAR on the whole, in terms of both data reduction and reduction quality. Further testing and experimentation is required - particularly in the area of attribute fuzzification. Work is also being carried out on a fuzzified dependency function. Ordinarily, the dependency function returns values for sets of attributes in the range [0,1]; the fuzzy function returns qualitative fuzzy labels for use in a new QUICKREDUCT algorithm. Additionally, research is being carried out into the potential utility of *fuzzy reducts*.

Acknowledgements

References

[1] Advanced Knowledge Technologies homepage: http://www.aktors.org/

[2] D. Dubois and H. Prade. Putting rough sets and fuzzy sets together. In R. Slowinski (Ed.) Intelligent Decision Support. Kluwer Academic Publishers, Dordrecht, 203-232. 1992.

[3] U. Höhle. Quotients with respect to similarity relations. Fuzzy Sets and Systems, 27:31-44. 1988.

[4] R. Jensen and Q. Shen. A Rough Set-Aided System for Sorting WWW Bookmarks. In N. Zhong et al. (Eds.), Web Intelligence: Research and Development. pp. 95-105, 2001.

[5] A. Lozowski, T.J. Cholewo and J.M. Zurada (1996). Crisp rule extraction from perceptron network classifiers. Proceedings of International Conference on Neural Networks, volume of Plenary, Panel and Special Sessions, 94 99.

[6] C.L. Blake & C.J. Merz (1998). UCI Repository of machine learning databases. Irvine, University of California, http://www.ics.uci.edu/~mlearn/MLRepository.html.

[7] Z. Pawlak. Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishing, Dordrecht, 1991.

[8] S.K. Pal and A. Skowron (Eds.). Rough-Fuzzy Hybridization : A New Trend in Decision Making. Springer Verlag, Singapore, 1999.

[9] Q. Shen and A. Chouchoulas. A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems. Engineering Applications of Artificial Intelligence, 13(3):263-278, 2000.