1 Introduction

Fuzzy-rough feature selection (FRFS) provides a means by which discrete or real-valued noisy data (or a mixture of both) can be effectively reduced without the need for user-supplied information. Additionally, this technique can be applied to data with continuous or nominal decision attributes, and as such can be applied to regression as well as classification datasets.

2 Fuzzy-Rough Sets

There have been two main lines of thought in the hybridization of fuzzy and rough sets, the constructive approach and the axiomatic approach. A general framework for the study of fuzzy-rough sets from both of these viewpoints is presented in [43]. For the constructive approach, generalized lower and upper approximations are defined based on fuzzy relations. Initially, these were fuzzy similarity/equivalence relations [9] but have since been extended to arbitrary fuzzy relations [43]. The axiomatic approach is primarily for the study of the mathematical properties of fuzzy-rough sets [37]. Here, various classes of fuzzy-rough approximation operators are characterized by different sets of axioms that guarantee the existence of types of fuzzy relations producing the same operators.

An original definition for fuzzy P-lower and P-upper approximations was given as follows [9]:

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

$$\mu_{\overline{P}X}(F_i) = \sup_x \min\{\mu_{F_i}(x), \mu_X(x)\} \quad \forall i \tag{2}$$

where F_i is a fuzzy equivalence class and X is the (fuzzy) concept to be approximated. The tuple $\langle \underline{P}X, \overline{P}X \rangle$ is called a fuzzy-rough set. These definitions diverge a little from the crisp upper and lower approximations, as the memberships of individual objects to the approximations are not explicitly available. As a result of this, the fuzzy lower and upper approximations are redefined as [12]:

$$\mu_{\underline{P}X}(x) = \sup_{F \in \mathbb{U}/P} \min(\mu_F(x), \inf_{y \in \mathbb{U}} \max\{1 - \mu_F(y), \mu_X(y)\}) \tag{3}$$

$$\mu_{\overline{P}X}(x) = \sup_{F \in \mathbb{U}/P} \min(\mu_F(x), \sup_{y \in \mathbb{U}} \min\{\mu_F(y), \mu_X(y)\}) \tag{4}$$

It can be seen that these definitions degenerate to traditional rough sets when all equivalence classes are crisp [11].

Also defined in the literature are rough-fuzzy sets [9], which can be seen to be a particular case of fuzzy-rough sets. A rough-fuzzy set is a generalization of a rough set, derived from the approximation of a fuzzy set in a crisp approximation space. In [39] it is argued that, to be consistent, the approximation of a crisp set in a fuzzy approximation space should be called a fuzzy-rough set, and the approximation of a fuzzy set in a crisp approximation space should be called a rough-fuzzy set, making the two models complementary. In this framework, the approximation of a fuzzy set in a fuzzy approximation space is considered to be a more general model, unifying the two theories. However, most researchers consider the traditional definition of fuzzy-rough sets in [9] as standard.

The specific use of min and max operators in the definitions above is expanded in [24], where a broad family of fuzzy-rough sets is constructed, each member represented by a particular implicator and t-norm. The properties of three well-known implicators (S-, R- and QL-implicators) are investigated. Further investigations in this area can be found in [8, 30, 38, 43].

2.0.1 Fuzzy-Rough QuickReduct

FRQuickReduct($\mathbb{C}$,$\mathbb{D}$).
$\mathbb{C}$, the set of all conditional attributes;
$\mathbb{D}$, the set of decision attributes.

$R \leftarrow \{\}$; $\gamma'_{best} = 0$; $\gamma'_{prev} = 0$
do
  $T \leftarrow R$
  $\gamma'_{prev} = \gamma'_{best}$
  foreach $x \in (\mathbb{C} - R)$
    if $\gamma'_{R\cup\{x\}}(\mathbb{D}) > \gamma'_T(\mathbb{D})$
      $T \leftarrow R \cup \{x\}$
      $\gamma'_{best} = \gamma'_T(\mathbb{D})$
  $R \leftarrow T$
until $\gamma'_{best} == \gamma'_{prev}$
return R

Figure 1: The fuzzy-rough QuickReduct algorithm

With these issues in mind, a fuzzy-rough hill-climbing search algorithm has been developed as given in Fig. 1. It employs the fuzzy-rough dependency function γ' to choose which attributes to add to the current reduct candidate in a manner similar to QuickReduct. The algorithm terminates when the addition of any remaining attribute does not increase the dependency (such a criterion could be used with the QuickReduct algorithm). As this fuzzy-rough degree of dependency measure is non-monotonic, it is possible that the hill-climbing search terminates having reached only a local optimum. The global optimum may lie elsewhere in the search space. As with the original QuickReduct algorithm, the algorithm may return a super-reduct (i.e. a reduct containing superfluous features) due to the non-optimality of the search heuristic used [41].

Note that with the fuzzy-rough QuickReduct algorithm, for a dimensionality of n, $(n^2+n)/2$ evaluations of the dependency function may be performed for the worst-case dataset. However, as FRFS is used for dimensionality reduction prior to any involvement of the system which will employ those attributes

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

belonging to the resultant reduct, this operation has no negative impact upon the run-time efficiency of the system.

2.0.2 Example

3 New Fuzzy Rough Feature Selection

This section presents three new techniques for fuzzy-rough feature selection, based on fuzzy similarity relations.

3.1 Fuzzy Lower Approximation-based FS

The previous method for fuzzy-rough feature selection used a fuzzy partitioning of the input space in order to determine fuzzy equivalence classes. Alternative definitions for the fuzzy lower and upper approximations can be found in [24], where a T-transitive fuzzy similarity relation is used to approximate a fuzzy concept X:

$$\mu_{\underline{R_P}X}(x) = \inf_{y \in \mathbb{U}} I(\mu_{R_P}(x,y), \mu_X(y)) \tag{5}$$

$$\mu_{\overline{R_P}X}(x) = \sup_{y \in \mathbb{U}} T(\mu_{R_P}(x,y), \mu_X(y)) \tag{6}$$

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

$$\mu_{R_P}(x,y) = \bigcap_{a \in P} \{\mu_{R_a}(x,y)\} \tag{7}$$

$\mu_{R_a}(x,y)$ is the degree to which objects x and y are similar for feature a. Many fuzzy similarity relations can be constructed for this purpose, for example:

$$\begin{aligned}
\mu_{R_a}(x,y) &= 1 - \frac{|a(x) - a(y)|}{|a_{max} - a_{min}|} && (8)\\
\mu_{R_a}(x,y) &= \exp\left(-\frac{(a(x) - a(y))^2}{2{\sigma_a}^2}\right) && (9)\\
\mu_{R_a}(x,y) &= \max(\min(\frac{(a(y) - (a(x) - \sigma_a))}{(a(x) - (a(x) - \sigma_a))}, \\
&\quad \frac{((a(x) + \sigma_a) - a(y))}{((a(x) + \sigma_a) - a(x))}, 0) && (10)
\end{aligned}$$

where ${\sigma_a}^2$ is the variance of feature a. As these relations do not necessarily display T-transitivity, the fuzzy transitive closure must be computed for each attribute [8]. The combination of feature relations in equation (7) has been shown to preserve T-transitivity [32].

3.1.1 Reduction

In a similar way to the original FRFS approach, the fuzzy positive region can be defined as:

$$\mu_{POS_{R_P}(Q)}(x) = \sup_{X \in \mathbb{U}/Q} \mu_{\underline{R_P}X}(x) \tag{11}$$

The resulting degree of dependency is:

$$\gamma'_P(Q) = \frac{\sum\limits_{x \in \mathbb{U}} \mu_{POS_{R_P}(Q)}(x)}{|\mathbb{U}|} \tag{12}$$

A fuzzy-rough reduct R can be defined as a 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 new fuzzy-rough QuickReduct algorithm can be constructed that operates in the same way as Fig. 1, but uses equation (12) to gauge subset quality. A proof of the monotonicity of the dependency function can be found in the paper [14]. Core features may be determined by considering the change in dependency of the full set of conditional features when individual attributes are removed:

$$Core(\mathbb{C}) = \{a \in \mathbb{C} | \gamma'_{\mathbb{C}-\{a\}}(Q) < \gamma'_{\mathbb{C}}(Q)\} \tag{13}$$

3.1.2 Example

The fuzzy connectives chosen for this example (and all others in this section) are the Łukasiewicz t-norm ($\max(x+y-1, 0)$) and the Łukasiewicz fuzzy implicator ($\min(1-x+y, 1)$). As recommended in [8], the Łukasiewicz t-norm is used as this produces fuzzy T-equivalence relations dual to that of a pseudo-metric. The use of the Łukasiewicz fuzzy implicator is also recommended as it is both a residual and S-implicator.

Using the fuzzy similarity measure defined in (10), the resulting relations are as follows for each feature in the dataset:

$$R_a(x,y) = \begin{pmatrix} 1.0 & 1.0 & 0.699 & 0.0 & 0.0 & 0.0 \\ 1.0 & 1.0 & 0.699 & 0.0 & 0.0 & 0.0 \\ 0.699 & 0.699 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.699 & 0.699 \\ 0.0 & 0.0 & 0.0 & 0.699 & 1.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.699 & 1.0 & 1.0 \end{pmatrix}$$

$$R_b(x,y) = \begin{pmatrix} 1.0 & 0.0 & 0.568 & 1.0 & 1.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.137 \\ 0.568 & 0.0 & 1.0 & 0.568 & 0.568 & 0.0 \\ 1.0 & 0.0 & 0.568 & 1.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.568 & 1.0 & 1.0 & 0.0 \\ 0.0 & 0.137 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

$$R_c(x,y) = \begin{pmatrix} 1.0 & 0.0 & 0.036 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.036 & 0.518 & 0.518 & 0.518 \\ 0.036 & 0.036 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.518 & 0.0 & 1.0 & 1.0 & 1.0 \\ 0.0 & 0.518 & 0.0 & 1.0 & 1.0 & 1.0 \\ 0.0 & 0.518 & 0.0 & 1.0 & 1.0 & 1.0 \end{pmatrix}$$

Again, the first step is to compute the lower approximations of each concept for each feature. Considering feature a and the decision concept {1,3,6} in the example dataset:

$$\mu_{\underline{R_a}\{1,3,6\}}(x) = \inf_{y \in \mathbb{U}} I(\mu_{R_a}(x,y), \mu_{\{1,3,6\}}(y))$$

For object 3, this is

$$\begin{aligned} \mu_{\underline{R_a}\{1,3,6\}}(3) &= \inf_{y \in \mathbb{U}} I(\mu_{R_a}(3,y), \mu_{\{1,3,6\}}(y)) \\ &= \inf\{I(0.699,1), I(0.699,0), I(1,1), \\ &\qquad I(0,0), I(0,0), I(0,1)\} \\ &= 0.301 \end{aligned}$$

For the remaining objects, this is:

$$\begin{aligned} \mu_{\underline{R_a}\{1,3,6\}}(1) &= 0.0 \\ \mu_{\underline{R_a}\{1,3,6\}}(2) &= 0.0 \\ \mu_{\underline{R_a}\{1,3,6\}}(4) &= 0.0 \\ \mu_{\underline{R_a}\{1,3,6\}}(5) &= 0.0 \\ \mu_{\underline{R_a}\{1,3,6\}}(6) &= 0.0 \end{aligned}$$

For concept $\{2, 4, 5\}$, the lower approximations are:

$$\begin{aligned} \mu_{\underline{R_a}\{2,4,5\}}(1) &= 0.0 \\ \mu_{\underline{R_a}\{2,4,5\}}(2) &= 0.0 \\ \mu_{\underline{R_a}\{2,4,5\}}(3) &= 0.0 \\ \mu_{\underline{R_a}\{2,4,5\}}(4) &= 0.301 \\ \mu_{\underline{R_a}\{2,4,5\}}(5) &= 0.0 \\ \mu_{\underline{R_a}\{2,4,5\}}(6) &= 0.0 \end{aligned}$$

Hence, the positive regions for each object are:

$$
\begin{aligned}
\mu_{POS_{R_a}(Q)}(1) &= 0.0 \\
\mu_{POS_{R_a}(Q)}(2) &= 0.0 \\
\mu_{POS_{R_a}(Q)}(3) &= 0.301 \\
\mu_{POS_{R_a}(Q)}(4) &= 0.301 \\
\mu_{POS_{R_a}(Q)}(5) &= 0.0 \\
\mu_{POS_{R_a}(Q)}(6) &= 0.0
\end{aligned}
$$

The resulting degree of dependency is therefore:

$$
\begin{aligned}
\gamma'_{\{a\}}(Q) &= \frac{\sum\limits_{x\in\mathbb{U}} \mu_{POS_{R_a}(Q)}(x)}{|\mathbb{U}|} \\
&= \frac{0.602}{6} \\
&= 0.1003
\end{aligned}
$$

Calculating the dependency degrees for the remaining features results in

$$\gamma'_{\{b\}}(Q) = 0.3597 \quad \gamma'_{\{c\}}(Q) = 0.4078$$

As feature c results in the largest increase in dependency degree, this feature is selected and added to the reduct candidate. The algorithm then evaluates the addition of all remaining features to this candidate. Fuzzy similarity relations are combined using (7). This produces the following evaluations:

$$\gamma'_{\{a,c\}}(Q) = 0.5501 \quad \gamma'_{\{b,c\}}(Q) = 1.0$$

Feature subset $\{b, c\}$ produces the maximum dependency value for this dataset, and the algorithm terminates. The dataset can now be reduced to these features only. The complexity of the algorithm is the same as that of FRFS in terms of the number of dependency evaluations. However, the explosive growth of the number of considered fuzzy equivalence classes is avoided through the use of fuzzy similarity relations and (7). This ensures that for one subset, only one fuzzy similarity relation is used to compute the fuzzy lower approximation.

3.2 Fuzzy Boundary Region-based FS

Most approaches to crisp rough set FS and all approaches to fuzzy-rough FS use only the lower approximation for the evaluation of feature subsets. The lower approximation contains information regarding the extent of certainty of object membership to a given concept. However, the upper approximation contains information regarding the degree of uncertainty of objects and hence this information can be used to discriminate between subsets. For example, two subsets may result in the same lower approximation but one subset may produce a smaller upper approximation. This subset will be more useful as there is less uncertainty concerning objects within the boundary region (the difference between upper and lower approximations). The fuzzy-rough boundary region for a fuzzy concept X may thus be defined:

$$\mu_{BND_{R_P}(X)}(x) = \mu_{\overline{R_P}X}(x) - \mu_{\underline{R_P}X}(x) \tag{14}$$

The fuzzy-rough negative region for all decision concepts can be defined as follows:

$$\mu_{NEG_{R_P}}(x) = N(\sup_{X \in \mathbb{U}/Q} \mu_{\overline{R_P}X}(x)) \tag{15}$$

In classical rough set theory, the negative region is always empty for partitions [42]. It is interesting to note that the fuzzy-rough negative region is also always empty when the decisions are crisp. However, this is not necessarily the case when decisions are fuzzy. Further details can be found in the paper [14].

3.2.1 Reduction

As the search for an optimal subset progresses, the object memberships to the boundary region for each concept diminishes until a minimum is achieved. For crisp rough set FS, the boundary region will be zero for each concept when a reduct is found. This may not necessarily be the case for fuzzy-rough FS due to the additional uncertainty involved. The uncertainty for a concept X using features in P can be calculated as follows:

$$U_P(X) = \frac{\sum_{x \in \mathbb{U}} \mu_{BND_{R_P}(X)}(x)}{|\mathbb{U}|} \tag{16}$$

This is the average extent to which objects belong to the fuzzy boundary region for the concept X. The total uncertainty degree for all concepts, given a feature subset P is defined as:

$$\lambda_P(Q) = \frac{\sum_{X \in \mathbb{U}/Q} U_P(X)}{|\mathbb{U}/Q|} \tag{17}$$

This is related to the conditional entropy measure which considers a combination of conditional probabilities $H(Q|P)$ in order to gauge the uncertainty present using features in P. In the crisp case, the minimization of this measure can be used to discover reducts: if the entropy for a feature subset P is zero, then the subset is a reduct [12].

Again, a QuickReduct-style algorithm can be constructed for locating fuzzy-rough reducts based on this measure. Instead of maximising the dependency degree, the task of the algorithm is to minimize the total uncertainty degree. When this reaches the minimum for the dataset, a fuzzy-rough reduct has been found. A proof of the monotonicity of the total uncertainty degree can be found in the paper [14].

3.2.2 Example

To determine the fuzzy boundary region, the lower and upper approximations of each concept for each feature must be calculated. Considering feature a and concept {1,3,6}:

$$\mu_{BND_{R_a}(\{1,3,6\})}(x) = \mu_{\overline{R_a}\{1,3,6\}}(x) - \mu_{\underline{R_a}\{1,3,6\}}(x)$$

For object 4, this is

$$\begin{aligned}
\mu_{BND_{R_a}(\{1,3,6\})}(4) &= \sup_{y\in\mathbb{U}} T(\mu_{R_a}(4,y), \mu_{\{1,3,6\}}(y)) \\
&\quad - \inf_{y\in\mathbb{U}} I(\mu_{R_a}(4,y), \mu_{\{1,3,6\}}(y)) \\
&= 0.699 - 0.0 \\
&= 0.699
\end{aligned}$$

For the remaining objects, this is:

$$\begin{aligned}
\mu_{BND_{R_a}(\{1,3,6\})}(1) &= 1.0 \\
\mu_{BND_{R_a}(\{1,3,6\})}(2) &= 1.0 \\
\mu_{BND_{R_a}(\{1,3,6\})}(3) &= 0.699 \\
\mu_{BND_{R_a}(\{1,3,6\})}(5) &= 1.0 \\
\mu_{BND_{R_a}(\{1,3,6\})}(6) &= 1.0
\end{aligned}$$

Hence, the uncertainty for concept {1,3,6} is:

$$\begin{aligned}
U_a(\{1,3,6\}) &= \frac{\sum_{x\in\mathbb{U}} \mu_{BND_{R_a}(\{1,3,6\})}(x)}{|\mathbb{U}|} \\
&= \frac{1.0 + 1.0 + 0.699 + 0.699 + 1.0 + 1.0}{6} \\
&= 0.899
\end{aligned}$$

For concept $\{2,4,5\}$, the uncertainty is:

$$\begin{aligned}
U_a(\{2,4,5\}) &= \frac{\sum_{x\in\mathbb{U}} \mu_{BND_{R_a}(\{2,4,5\})}(x)}{|\mathbb{U}|} \\
&= \frac{1.0 + 1.0 + 0.699 + 0.699 + 1.0 + 1.0}{6} \\
&= 0.899
\end{aligned}$$

From this, the total uncertainty for feature a is calculated as follows:

$$\begin{aligned}
\lambda_a(Q) &= \frac{\sum_{X\in\mathbb{U}/Q} U_a(X)}{|\mathbb{U}/Q|} \\
&= \frac{0.899 + 0.899}{2} \\
&= 0.899
\end{aligned} \tag{18}$$

The values of the total uncertainty for the remaining features are:

$$\lambda_{\{b\}}(Q) = 0.640 \quad \lambda_{\{c\}}(Q) = 0.592$$

As feature c results in the smallest total uncertainty, it is chosen and added to the reduct candidate. The algorithm then considers the addition of the remaining features to the subset:

$$\lambda_{\{a,c\}}(Q) = 0.500 \quad \lambda_{\{b,c\}}(Q) = 0.0$$

The subset $\{b, c\}$ results in the minimal uncertainty for the dataset, and the algorithm terminates. This is the same subset as that chosen by the fuzzy lower approximation-based method above. Again, the complexity of the algorithm is the same as that of FRFS, but avoids the Cartesian product of fuzzy equivalence classes. However, for each evaluation, both the fuzzy lower and upper approximations are considered and hence the calculation of the fuzzy boundary region is more costly than that of the fuzzy lower approximation alone.

3.3 Fuzzy Discernibility Matrix-based FS

As mentioned previously, there are two main branches of research in crisp rough set-based FS: those based on the dependency degree and those based on discernibility matrices. The developments given above are solely concerned with the extension of the dependency degree to the fuzzy-rough case. Hence, methods constructed based on the crisp dependency degree can be employed for fuzzy-rough FS.

By extending the discernibility matrix to the fuzzy case, it is possible to employ approaches similar to those in crisp rough set FS to determine fuzzy-rough reducts. A first step toward this is presented in [31, 34] where a crisp discernibility matrix is constructed for fuzzy-rough selection. A threshold is used, breaking the rough set ideology, which determines which features are to appear in the matrix entries. However, information is lost in this process as membership degrees are not considered. Search based on the crisp discernibility may result in reducts that are not true fuzzy-rough reducts.

3.3.1 Fuzzy Discernibility

The approach presented here extends the crisp discernibility matrix by employing fuzzy clauses. Each entry in the fuzzy discernibility matrix is a fuzzy set, to which every feature belongs to a certain degree. The extent to which a feature a belongs to the fuzzy clause C_{ij} is determined by the fuzzy discernibility measure:

$$\mu_{C_{ij}}(a) = N(\mu_{R_a}(i, j)) \tag{19}$$

where N denotes fuzzy negation and $\mu_{R_a}(i, j)$ is the fuzzy similarity of objects i and j, and hence $\mu_{C_{ij}}(a)$ is a measure of the fuzzy discernibility. For the crisp case, if $\mu_{C_{ij}}(a) = 1$ then the two objects are distinct for this feature; if $\mu_{C_{ij}}(a)$

= 0, the two objects are identical. For fuzzy cases where $\mu_{C_{ij}}(a) \in (0,1)$, the objects are partly discernible. (The choice of fuzzy similarity relation must be identical to that of the fuzzy-rough dependency degree approach to find corresponding reducts.) Each entry in the fuzzy indiscernibility matrix is then a set of attributes and their corresponding memberships:

$$C_{ij} = \{a_x | a \in \mathbb{C}, x = N(\mu_{R_a}(i,j))\} \;\; i,j = 1,...,|\mathbb{U}| \tag{20}$$

For example, an entry C_{ij} in the fuzzy discernibility matrix might be:

$$C_{ij} : \{a_{0.4}, b_{0.8}, c_{0.2}, d_{0.0}\}$$

This denotes that $\mu_{C_{ij}}(a) = 0.4$, $\mu_{C_{ij}}(b) = 0.8$, etc. In crisp discernibility matrices, these values are either 0 or 1 as the underlying relation is an equivalence relation. The example clause can be viewed as indicating the value of each feature - the extent to which the feature discriminates between the two objects i and j. The core of the dataset is defined as:

$$\begin{aligned} Core(\mathbb{C}) \;\; = \;\; & \{a \in \mathbb{C} | \exists C_{ij}, \mu_{C_{ij}}(a) > 0, \\ & \forall f \in \{\mathbb{C} - a\} \mu_{C_{ij}}(f) = 0\} \end{aligned} \tag{21}$$

3.3.2 Fuzzy Discernibility Function

As with the crisp approach, the entries in the matrix can be used to construct the fuzzy discernibility function:

$$f_D(a_1^*, ..., a_m^*) = \wedge\{\vee\, C_{ij}^* | 1 \leq j < i \leq |\mathbb{U}|\} \tag{22}$$

where $C_{ij}^* = \{a_x^* | a_x \in C_{ij}\}$. The function returns values in $[0,1]$, which can be seen to be a measure of the extent to which the function is satisfied for a given assignment of truth values to variables. To discover reducts from the fuzzy discernibility function, the task is to find the minimal assignment of the value 1 to the variables such that the formula is maximally satisfied. By setting all variables to 1, the maximal value for the function can be obtained as this provides the most discernibility between objects.

Crisp discernibility matrices can be simplified by removing duplicate entries and clauses that are supersets of others. A similar degree of simplification can be achieved for fuzzy discernibility matrices. Duplicate clauses can be removed as a subset that satisfies one clause to a certain degree will always satisfy the other to the same degree.

3.3.3 Decision-relative Fuzzy Discernibility Matrix

As with the crisp discernibility matrix, for a decision system the decision feature must be taken into account for achieving reductions; only those clauses with different decision values are included in the crisp discernibility matrix. For the fuzzy version, this is encoded as:

$$\begin{aligned} f_D(a_1^*, ..., a_m^*) &= \{\wedge\{\{\vee C_{ij}^*\} \leftarrow q_{N(\mu_{R_q}(i,j))}\}| \\ & \quad 1 \leq j < i \leq |\mathbb{U}|\} \end{aligned} \tag{23}$$

for decision feature q, where $\leftarrow$ denotes fuzzy implication. This construction allows the extent to which decision values differ to affect the overall satisfiability of the clause. If $\mu_{C_{ij}}(q) = 1$ then this clause provides maximum discernibility (i.e. the two objects are maximally different according to the fuzzy similarity measure). When the decision is crisp and crisp equivalence is used, $\mu_{C_{ij}}(q)$ becomes 0 or 1.

3.3.4 Reduction

For the purposes of finding reducts, use of the fuzzy intersection of all clauses in the fuzzy discernibility function may not provide enough information for evaluating subsets. Here, it may be more informative to consider the individual satisfaction of each clause for a given set of features. The degree of satisfaction of a clause C_{ij} for a subset of features P is defined as:

$$SAT_P(C_{ij}) = \bigcup_{a \in P} \{\mu_{C_{ij}}(a)\} \tag{24}$$

Returning to the example, if the subset $P = \{a, c\}$ is chosen, the resulting degree of satisfaction of the clause is

$$SAT_P(C_{ij}) = \{0.4 \vee 0.2\} = 0.6$$

using the Łukasiewicz t-conorm, $\min(1, x + y)$.

For the decision-relative fuzzy indiscernibility matrix, the decision feature q must be taken into account also:

$$SAT_{P,q}(C_{ij}) = SAT_P(C_{ij}) \leftarrow \mu_{C_{ij}}(q) \tag{25}$$

For the example clause, if the corresponding decision values are crisp and are different, the degree of satisfaction of the clause is

$$\begin{aligned} SAT_{P,q}(C_{ij}) &= SAT_P(C_{ij}) \leftarrow 1 \\ &= 0.6 \leftarrow 1 \\ &= 0.6 \end{aligned}$$

For a subset P, the total satisfiability of all clauses can be calculated as

$$SAT(P) = \frac{\sum\limits_{i,j \in \mathbb{U}, i \neq j} SAT_{P,q}(C_{ij})}{\sum\limits_{i,j \in \mathbb{U}, i \neq j} SAT_{\mathbb{C},q}(C_{ij})} \tag{26}$$

where $\mathbb{C}$ is the full set of conditional attributes, and hence the denominator is a normalizing factor. If this value reaches 1 for a subset P, then the subset is a fuzzy-rough reduct. A proof of the monotonicity of the function $SAT(P)$ can be found in the paper [14].

Many methods available from the literature for the purpose of finding reducts for crisp discernibility matrices are applicable here also. The Johnson Reducer [19] is extended and used herein to illustrate the concepts involved. This is a simple greedy heuristic algorithm that is often applied to discernibility functions to find a single reduct. Subsets of features found by this process have no guarantee of minimality, but are generally of a size close to the minimal.

The algorithm begins by setting the current reduct candidate, P, to the empty set. Then, each conditional feature appearing in the discernibility function is evaluated according to the heuristic measure used. For the standard Johnson algorithm, this is typically a count of the number of appearances a feature makes within clauses; features that appear more frequently are considered to be more significant. The feature with the highest heuristic value is added to the reduct candidate and all clauses in the discernibility function containing this feature are removed. As soon as all clauses have been removed, the algorithm terminates and returns the subset P. P is assured to be a fuzzy-rough reduct as all clauses contained within the discernibility function have been addressed. However, as with the other approaches, the subset may not necessarily have minimal cardinality.

The complexity of the algorithm is the same as that of FRFS in that $O((n^2+n)/2)$ calculations of the evaluation function ($SAT(P)$) are performed in the worst case. Additionally, this approach requires the construction of the fuzzy discernibility matrix, which has a complexity of $O(a{*}o^2)$ for a dataset containing a attributes and o objects.

3.3.5 Example

For the example dataset, the fuzzy discernibility matrix needs to be constructed based on the fuzzy discernibility given in equation (19) using the standard negator, and fuzzy similarity in equation (10). For objects 2 and 3, the resulting fuzzy clause is:

$$\{a_{0.301} \vee b_{1.0} \vee c_{0.964}\} \leftarrow q_{1.0}$$

where $\leftarrow$ denotes fuzzy implication. The fuzzy discernibility of objects 2 and 3 for attribute a is 0.301, indicating that the objects are partly discernible for this feature. The objects are fully discernible with respect to the decision feature, indicated by $q_{1.0}$. The full set of clauses is:

$$
\begin{array}{llll}
C_{12}: & \{a_{0.0} \vee b_{1.0} \vee c_{1.0}\} & \leftarrow & q_{1.0} \\
C_{13}: & \{a_{0.301} \vee b_{0.432} \vee c_{0.964}\} & \leftarrow & q_{0.0} \\
C_{14}: & \{a_{1.0} \vee b_{0.0} \vee c_{1.0}\} & \leftarrow & q_{1.0} \\
C_{15}: & \{a_{1.0} \vee b_{0.0} \vee c_{1.0}\} & \leftarrow & q_{1.0} \\
C_{16}: & \{a_{1.0} \vee b_{1.0} \vee c_{1.0}\} & \leftarrow & q_{0.0} \\
C_{23}: & \{a_{0.301} \vee b_{1.0} \vee c_{0.964}\} & \leftarrow & q_{1.0} \\
C_{24}: & \{a_{1.0} \vee b_{1.0} \vee c_{0.482}\} & \leftarrow & q_{0.0} \\
C_{25}: & \{a_{1.0} \vee b_{1.0} \vee c_{0.482}\} & \leftarrow & q_{0.0} \\
C_{26}: & \{a_{1.0} \vee b_{0.863} \vee c_{0.482}\} & \leftarrow & q_{1.0} \\
C_{34}: & \{a_{1.0} \vee b_{0.431} \vee c_{1.0}\} & \leftarrow & q_{1.0} \\
C_{35}: & \{a_{1.0} \vee b_{0.431} \vee c_{1.0}\} & \leftarrow & q_{1.0} \\
C_{36}: & \{a_{1.0} \vee b_{1.0} \vee c_{1.0}\} & \leftarrow & q_{0.0} \\
C_{45}: & \{a_{0.301} \vee b_{0.0} \vee c_{0.0}\} & \leftarrow & q_{0.0} \\
C_{46}: & \{a_{0.301} \vee b_{1.0} \vee c_{0.0}\} & \leftarrow & q_{1.0} \\
C_{56}: & \{a_{0.0} \vee b_{1.0} \vee c_{0.0}\} & \leftarrow & q_{1.0}
\end{array}
$$

The feature selection algorithm then proceeds in the following way. Each individual feature is evaluated according to the measure defined in equation (26). For feature a, this is:

$$
\begin{aligned}
SAT(\{a\}) &= \frac{\sum_{i,j \in \mathbb{U}, i \neq j} SAT_{\{a\},q}(C_{ij})}{\sum_{i,j \in \mathbb{U}, i \neq j} SAT_{\mathbb{C},q}(C_{ij})} \\
&= \frac{11.601}{15} \\
&= 0.773
\end{aligned}
$$

Similarly for the remaining features:

$$SAT(\{b\}) = 0.782 \quad SAT(\{c\}) = 0.830$$

The feature that produces the largest increase in satisfiability is c. This feature is added to the reduct candidate, and the search continues:

$$SAT(\{a, c\}) = 0.887 \quad SAT(\{b, c\}) = 1.0$$

The subset $\{b, c\}$ is found to satisfy all clauses maximally, and the algorithm terminates. This subset is a fuzzy-rough reduct.

4 Vaguely Quantified Rough Sets

References

[1] "Rough Sets and Current Trends in Computing," *Proc. Third Int'l Conf.*, J.J. Alpigini, J.F. Peters, J. Skowronek, and N. Zhong, eds., 2002.

[2] R.B. Bhatt and M. Gopal, "On fuzzy-rough sets approach to feature selection," *Pattern Recognition Letters*, vol. 26, no. 7, pp. 965–975, 2005.

[3] R.B. Bhatt and M. Gopal, "On the compact computational domain of fuzzy-rough sets," *Pattern Recognition Letters*, vol. 26, no. 11, pp. 1632–1640, 2005.

[4] C.L. Blake and C.J. Merz. UCI Repository of machine learning databases. Irvine, University of California, 1998. http://www.ics.uci.edu/~mlearn/.

[5] A. Chouchoulas and Q. Shen, "Rough Set-Aided Keyword Reduction for Text Categorisation," *Applied Artificial Intelligence*, vol. 15, no. 9, pp. 843–873, 2001.

[6] W.W. Cohen, "Fast effective rule induction," In *Proceedings of the 12th International Conference on Machine Learning*, pp. 115–123, 1995.

[7] M. Dash and H. Liu, "Feature Selection for Classification," *Intelligent Data Analysis*, vol. 1, no. 3, 1997.

[8] M. De Cock, C. Cornelis, and E.E. Kerre, "Fuzzy Rough Sets: The Forgotten Step," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 121–130, 2007.

[9] D. Dubois and H. Prade, "Putting Rough Sets and Fuzzy Sets Together," *Intelligent Decision Support*, pp. 203–232, 1992.

[10] I. Düntsch and G. Gediga, *Rough Set Data Analysis: A Road to Non-Invasive Knowledge Discovery*. Bangor: Methodos, 2000.

[11] R. Jensen and Q. Shen, "Fuzzy-Rough Attribute Reduction with Application to Web Categorization," *Fuzzy Sets and Systems*, vol. 141, no. 3, pp. 469–485, 2004.

[12] R. Jensen and Q. Shen, "Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough Based Approaches," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 12, pp. 1457–1471, 2004.

[13] R. Jensen and Q. Shen, "Fuzzy-Rough Sets Assisted Attribute Selection," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 73–89, 2007.

[14] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," Accepted for publication in: *IEEE Transactions on Fuzzy Systems*.

[15] J. Komorowski, Z. Pawlak, L. Polkowski and A. Skowron, "Rough Sets: A Tutorial," In [20], pp. 3-98. 1999.

[16] P. Langley, "Selection of Relevant Features in Machine Learning," *Proc. AAAI Fall Symp. Relevance*, pp. 1–5, 1994.

[17] P. Lingras and R. Jensen, "Survey of Rough and Fuzzy Hybridization," *Proceedings of the 16th International Conference on Fuzzy Systems* (FUZZ-IEEE'07), pp. 125–130, 2007.

[18] H. S. Nguyen and A. Skowron, "Boolean Reasoning for Feature Extraction Problems," ISMIS pp. 117-126, 1997.

[19] A. Øhrn, "Discernibility and Rough Sets in Medicine: Tools and Applications," Department of Computer and Information Science. Trondheim, Norway, Norwegian University of Science and Technology: 239. 1999.

[20] *Rough-Fuzzy Hybridization: A New Trend in Decision Making*, S.K. Pal and A. Skowron, eds. Springer Verlag, 1999.

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

[22] "Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems," *Studies in Fuzziness and Soft Computing*, L. Polkowski, T.Y. Lin, and S. Tsumoto, eds., vol. 56, Physica-Verlag, 2000.

[23] L. Polkowski, "Rough Sets: Mathematical Foundations," *Advances in Soft Computing*, Physica Verlag, 2002.

[24] A.M. Radzikowska and E.E. Kerre, "A comparative study of fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 126, no. 2, pp. 137–155, 2002.

[25] Q. Shen and A. Chouchoulas, "A Fuzzy-Rough Approach for Generating Classification Rules," *Pattern Recognition*, vol. 35, no. 11, pp. 341–354, 2002.

[26] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[27] A. Skowron and C. Rauszer, "The discernibility matrices and functions in Information Systems," In: [28], pp. 331-362. 1992.

[28] *Intelligent Decision Support*, R. Slowinski, ed., Kluwer Academic Publishers, 1992.

[29] R.W. Swiniarski and A. Skowron, "Rough Set Methods in Feature Selection and Recognition," *Pattern Recognition Letters*, vol. 24, no. 6, pp. 833–849, 2003.

[30] H. Thiele, "Fuzzy Rough Sets versus Rough Fuzzy Sets - An Interpretation and a Comparative Study Using Concepts of Modal Logics," Technical Report no. CI-30/98, Univ. of Dortmund, 1998.

[31] G.C.Y. Tsang, D. Chen, E.C.C. Tsang, J.W.T. Lee, and D.S. Yeung, "On attributes reduction with fuzzy rough sets," *Proc. 2005 IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2775–2780, 2005.

[32] M. Wallace, Y. Avrithis and S. Kollias, "Computationally efficient sup-t transitive closure for sparse fuzzy binary relations," *Fuzzy Sets and Systems*, vol. 157, no. 3, pp. 341–372, 2006.

[33] J. Wang and J. Wang, "Reduction Algorithms Based on Discernibility Matrix: The Ordered Attributes Method," *J. Comput. Sci. & Technol.*, vol. 16, no. 6, pp. 489–504, 2001.

[34] X.Z. Wang, Y. Ha, and D. Chen, "On the reduction of fuzzy rough sets," *Proc. 2005 International Conference on Machine Learning and Cybernetics*, vol. 5, pp. 3174–3178, 2005.

[35] I.H. Witten and E. Frank, "Generating Accurate Rule Sets Without Global Optimization," In *Proceedings of the 15th International Conference on Machine Learning*, Morgan Kaufmann Publishers, San Francisco, 1998.

[36] I.H. Witten and E. Frank, *Data Mining: Practical machine learning tools with Java implementations.* Morgan Kaufmann, San Francisco, 2000.

[37] W.Z. Wu and W.X. Zhang, "Constructive and axiomatic approaches of fuzzy approximation operators," *Information Sciences*, vol. 159, no.3-4, pp. 233–254, 2004.

[38] W.Z. Wu, Y. Leung, and J.S. Mi, "On characterizations of (I,T)-fuzzy rough approximation operators," *Fuzzy Sets and Systems*, vol. 154, no. 1, pp. 76–102, 2005.

[39] Y.Y. Yao, "Combination of rough and fuzzy sets based on α-level sets," in: T.Y. Lin, N. Cereone (Eds.), *Rough Sets and Data Mining: Analysis of Imprecise Data*, Kluwer Academic Publishers, pp. 301–321, 1997.

[40] Y.Y. Yao, "A Comparative Study of Fuzzy Sets and Rough Sets," *Information Sciences*, vol. 109, pp. 21–47, 1998.

[41] Y.Y. Yao, Y. Zhao and J. Wang, "On reduct construction algorithms," *Proceedings of the First International Conference on Rough Sets and Knowledge Technology* (RSKT06), 297–304, 2006.

[42] Y.Y. Yao, "Decision-theoretic rough set models," *Proceedings of the International Conference on Rough Sets and Knowledge Technology* (RSKT07), LNAI 4481, pp. 1–12, 2007.

[43] D.S. Yeung, D. Chen, E.C.C. Tsang, J.W.T. Lee, and W. Xizhao, "On the Generalization of Fuzzy Rough Sets," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 3, pp. 343–361, 2005.

[44] L.A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[45] L.A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," *Information Sciences*, vol. 8, pp. 199–249, 301–357; vol. 9: 43–80, 1975.