*The Knowledge Engineering Review*, Vol. 25:4, 365–395. © Cambridge University Press, 2010
doi:10.1017/S0269888910000263

# On rough sets, their recent extensions and applications

N. MAC PARTHALÁIN and Q. SHEN

*Department of Computer Science, Aberystwyth University, Aberystwyth, Ceredigion, SY23 3DB, Wales, UK; e-mail: ncm@aber.ac.uk, qqs@aber.ac.uk*

**Abstract**

Rough set theory (RST) has enjoyed an enormous amount of attention in recent years and has been applied to many real-world problems including data mining, pattern recognition, and intelligent control. Much research has recently been carried out in respect of both the development of the underlying theory and the application to new problem domains. This paper attempts to summarize the advances in RST, its extensions, and their applications. It also identifies important areas which require further investigation. Typical example application domains are examined which demonstrate the success of the application of RST to a wide variety of areas and disciplines, and which also exhibit the strengths and limitations of the respective underlying approaches.

## 1 Introduction

The ability to deal effectively with insufficient or imperfect knowledge is a central motivating factor in much of the research in the field of computational intelligence. In the areas of machine learning, data mining, pattern recognition, and intelligent control, the ability to handle such knowledge is of primary importance both in terms of theoretical advancement and practical applications. The work in the area of rough set theory (RST; Pawlak, 1982), Pawlak (1991) offers perhaps one of the most distinct and recent approaches in this respect.

Such is the worldwide nature of the attention that RST has attracted since its inception (Komorowski *et al*., 1999) that much research and development has been carried out not only in applying the theory to many and varied problem domains, but also to extending it theoretically. This has resulted in a significant breadth and depth of work in the area. RST (Pawlak, 1982) has been used as a tool to discover data dependencies and to reduce the number of attributes contained in a data set using the data alone, requiring no additional information (Pawlak, 1982; Pawlak, 1991; Polkowski & Skowron, 1998; Skowron *et al*., 2002). Since its inception, RST has been successfully utilized to devise mathematically sound and often computationally efficient techniques for addressing problems such as knowledge discovery from data, data reduction, data significance evaluation, decision rule generation, and data-driven inference interpretation (Pawlak, 2003). Given a data set with discretized attribute values, it is possible to find a subset (termed reduct) of the original attributes using RST that are most informative; all other attributes can be removed from the data set with minimal information loss. RST possesses many features in common (to a certain extent) with the Dempster–Shafer theory of evidence (Shafer, 1976) and fuzzy set theory (FST; Zadeh, 1965). It works by making use of the granular structure of the data only. This is a major difference when compared with the Dempster–Shafer theory and fuzzy set theory, which require probability assignments and membership values, respectively. The use of only the data and their granularity ensures that no other assumptions are made about the data. This approach has led to some researchers suggesting that this is a disadvantage rather than an

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

advantage of RST (Komorowski *et al*., 1999) as other numerical and contextual aspects are effectively ignored. However, in disregarding such supplemental information, model assumptions can be minimized.

Formally, a rough set is the approximation of a vague concept (set) by a pair of precise concepts, called lower and upper approximations, which are a classification of the domain of interest into disjoint categories. The lower approximation is a description of the domain objects which are known with certainty to belong to the concept of interest, whereas the upper approximation is a description of the objects which possibly belong to the concept. The approximations are constructed with regard to a particular subset of attributes or features.

One of the primary drawbacks of RST lies in its inability to deal with real world data. Owing mainly to the granular approach that RST uses to handle data, and the strict structure of equivalence imposed, it does not allow any flexibility when dealing with measurement noise or imperfection that is prevalent in real world data. However, most data sets contain real-valued features and so it becomes necessary to perform a discretization step before employing RST for knowledge discovery. Take for instance a weather forecasting system which records a number of meteorological attributes, with one in particular that might be *average rainfall*. In reality, this is a continuous and real-valued measurement. However, in order to apply RST to such a problem, this attribute must be discretized with a set of labels such as *light*, *medium*, and *heavy*. This imposes subjective human judgement on what is otherwise an objective measurement.

The deficiency of RST in handling real-valued data has resulted over the years in the development of a number of extensions which aim to address this problem. There are two areas of RST which have been considerably exploited in order to achieve this: modification of the equivalence relation, and manipulation of the subset operator. These are the primary operations of RST and it is unsurprising, therefore, that a number of extensions have been proposed with regard to these areas. The tolerance rough set model (TRSM; Skowron & Stepaniuk, 1994) is a typical example of an attempt to address this problem through the modification of the equivalence relation. Variable precision rough sets (VPRS; Ziarko, 1993) allow the relaxation of the subset operator of traditional RST. This approach was originally formulated to analyse and identify data patterns which represent statistical trends.

In addition to the use of alternative equivalence relations and modification of the subset operator, there is also a third aspect of RST which has been exploited, that of the use of the information contained in the boundary region, or region of uncertainty between the lower and upper approximations (Hu *et al*., 2007a; Mac Parthaláin *et al*., 2007). This information, although uncertain, can be useful in maximizing the performance of RST without changing the underlying model or modifying the subset operators.

As well as directly extending RST, it has also been hybridized with other soft computing methods such as fuzzy sets (Zadeh, 1965), genetic algorithms (GAs), neural networks, and statistical methods such as principal component analysis (PCA; Devijver & Kittler, 1982), etc. Such hybridization has highlighted the value of employing RST, as its use often results in methods which outperform such methods individually. In particular, the hybridization of RST with FST (Zadeh, 1965) to form fuzzy-RST (Dubois & Prade, 1992) is perhaps the most important of all. Fuzzy-RST (Dubois & Prade, 1992) attempts to take advantage of the complementary nature of fuzzy sets and rough sets. The significance of this work is reflected in the level of research carried out in this area and also to the number of applications of fuzzy-RST.

This paper attempts to offer a brief overview of the basic concepts which underpin RST. In particular, the more recent extensions of RST are examined, as well as a look at some representative theoretical application areas such as classification, clustering, and feature selection. These theoretical applications are supported by three successful practical application examples in breast cancer risk assessment, document classification, and gene expression, respectively.

The remainder of this paper is organized as follows. In Section 2, the preliminary concepts and theoretical foundation of RST are outlined. Various rough set extensions (both past and recent) such as tolerance rough sets, VPRS, dominance-based rough sets, vaguely quantified rough sets,

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

and others are examined in Section 3. The hybridization of RST with other techniques is discussed in Section 4, with particular emphasis on fuzzy-rough sets. A range of both theoretical and real world example applications with regard to RST, and the above-mentioned extensions are discussed in Section 5. The final section concludes the paper and discusses identified important further work.

## 2 Rough sets

In this section, the basic notions, definitions, and operations of RST are described. The upper and lower approximation concepts, as well as how these can be used to minimize data, are also explored. A small example is used to demonstrate all of the concepts described and show the individual steps involved in employing RST. Heuristics for discovering reducts, and search techniques are also discussed.

### 2.1 Basic concepts and theoretical background

Central to RST is the concept of indiscernibility. 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. For any $P \subseteq \mathbb{A}$, there is an associated equivalence relation $IND(P)$:

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

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

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

*where*,

$$\mathbb{U}/IND(\{a\}) = \{\{x \mid a(x) = b,\ x \in \mathbb{U}\} \mid b \in V_a\} \tag{3}$$

*and*,

$$A \otimes B = \{X \cap Y : \forall X \in A, \forall Y \in B, X \cap Y \neq \theta\} \tag{4}$$

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 by $[x]_P$.

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 $X$:

$$\underline{P}X = \{x \mid [x]_P \subseteq X\} \tag{5}$$

$$\overline{P}X = \{x \mid [x]_P \cap X \neq \theta\} \tag{6}$$

Let $P$ and $Q$ be equivalence relations over $\mathbb{U}$; then the positive, negative, and boundary regions are defined by:

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

$$NEG_P(Q) = \mathbb{U} - \bigcup_{X \in \mathbb{U}/Q} \overline{P}X \tag{8}$$

$$BND_P(Q) = \bigcup_{X \in \mathbb{U}/Q} \overline{P}X - \bigcup_{X \in \mathbb{U}/Q} \underline{P}X \tag{9}$$

The positive region contains all objects of $\mathbb{U}$ that can be classified to classes of $\mathbb{U}/Q$ using the information in attribute $P$. The boundary region, $BND_P(Q)$, is the set of objects that can possibly,

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

**Table 1** Example data set

| $x \in U$ | $a$ | $b$ | $c$ | $d$ | → | $e$ |
|---|---|---|---|---|---|---|
| 1 | M | L | N | N | | H |
| 2 | L | M | M | M | | F |
| 3 | M | M | L | N | | F |
| 4 | M | L | N | L | | G |
| 5 | N | N | L | M | | G |
| 6 | N | M | M | M | | F |
| 7 | L | M | M | L | | G |

but not certainly, be classified in this way. The negative region, $NEG_P(Q)$, is the set of objects that cannot be classified to classes of $\mathbb{U}/Q$.

*2.1.1 Example*

To illustrate the above concepts, a short example in the form of an information system is employed. There are four conditional attributes: $a$, $b$, $c$, and $d$, and a single decisional attribute, $e$.

Using the indiscernibilty concept, the data in Table 1 can be partitioned according to the outcome. $V_a$ is the set of values that attribute $a$ may take (in this case $L$, $M$, or $N$). In a decision system, $A = \{\mathbb{C} \cup \mathbb{D}\}$ where $\mathbb{C}$ denotes the set of condition attributes and $\mathbb{D}$ denotes the set of decision attribute(s). There are associated equivalence relations with any $P \subseteq \mathbb{A}$:

$$IND(P) = \{(x, y) \in \mathbb{U}^2 \mid \forall a \in P, a(x) = a(y)\} \tag{10}$$

For the data in Table 1—the partition of $\mathbb{U}$ by the attribute $a$ would be:

$$\mathbb{U}/IND(\{a\}) = \{\{1, 3, 4\}, \{2, 7\}, \{5, 6\}\} \tag{11}$$

And for the same table using attributes $\{b,c\}$

$$\mathbb{U}/IND(\{b, c\}) = \{\{1, 4\}, \{2, 6, 7\}, \{3\}, \{5\}\} \tag{12}$$

This relates to the partition or grouping of the attributes where: $a = L$ (objects 1, 3, and 4), $a = M$ (objects 2 and 7), and $a = N$ (objects 5 and 6). The equivalence classes of the $P$-indiscernibility relation are denoted by $[x]_P$. Let $\mathbb{X} \subseteq \mathbb{U}$. $X$ can be approximated using only the information within $P$ by formulating lower and upper approximations of $X$ as described previously.

*2.2 Rough set dependency and other measures*

An important aspect of data analysis is the discovery of dependencies between attributes. From an intuitive point of view, an attribute or a set of attributes $Q$ can depend on a set of attributes $P$, denoted by $P \Rightarrow Q$ if all values of attribute(s) in $Q$ are determined uniquely by values of attribute(s) from $P$. Another way of describing this is that $Q$ depends totally on $P$ if a functional dependency exists between the values of $Q$ and $P$.

Referring to the example in the previous section, the rough set dependency of the set of attributes $Q$ on a set of attributes $P$ can be seen. For $P, Q \subset \mathbb{A}$, it can be said that $Q$ depends on $P$ to a degree $k$ (where $k \in [0,1]$) denoted by $P \Rightarrow_k Q$ if:

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|\mathbb{U}|} \tag{13}$$

*where*

$$POS_P(Q) = \bigcup_{X \in \mathbb{U}/Q} \underline{P}(X) \tag{14}$$

is the positive region of the partition of the universe with respect to $P$ (i.e. the set of all elements that can be classified uniquely into sets of the partition $\mathbb{U}/Q$ in terms of $P$).

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

If $k = 1$, $Q$ is completely dependent on $P$, if $k < 1$ $Q$ is partially dependent (to a degree—$k$) on $P$ and obviously, if $k = 0$, $Q$ is completely non-dependent on $P$. Calculation of the relevant dependencies of each attribute (or group of attributes) allows the significance of that attribute (or group) to be realized.

Taking the data from the example decision table (Table 1), the degree of dependency of attribute $\{e\}$ upon the attributes $\{b, c\}$ is:

$$\gamma_{\{b,c\}}(\{e\}) = \frac{|POS_{\{b,c\}}(\{e\})|}{|\mathbb{U}|} = \frac{|\{3,5\}|}{|\{1,2,3,4,5,6,7,\}|} = \frac{2}{7}$$

For the application of feature selection, the minimization of attributes can be realized through the comparison of equivalence relations generated by sets of attributes ($\{b, c\}$ for the purpose of the previous example). Attributes are removed such that the minimized set provides an equivalent predictive characteristic as the initial decision feature. This minimized set is termed a reduct and can be defined as a subset $R$ of the conditional attribute set $\mathbb{C}$ such that $\gamma_R(\mathbb{D}) = \gamma_{\mathbb{C}}(\mathbb{D})$.

Other measures have also been used to discover rough set reducts. For instance, in Han *et al.* (2004), a feature selection method which is based on an alternative dependency measure is presented. This technique was proposed to avoid the expensive calculation of discernibility functions or positive regions. The authors replace the traditional rough set dependency measure with the relative dependency measure, defined as follows for an attribute subset $P$:

$$\kappa_P(\mathbb{D}) = \frac{|\mathbb{U}/IND(P)|}{|\mathbb{U}/IND(P \cup \mathbb{D})|} \tag{15}$$

The authors then demonstrate that $R$ is a reduct if and only if $\kappa_R(D) = \kappa_C(D)$ and that $\forall X \subset R$, $\kappa_X(D) \neq \kappa_C(D)$.

In addition, the entropy measure has been used in Jensen and Shen (2004b) to discover smaller reducts than the rough set dependency measure alone. In this approach, although entropy is used in the search for reducts, rough set dependency is still used as a termination criterion.

### 2.3 Minimal reducts and reduct discovery

A method for reducing data, demonstrated in the previous example, identifies equivalence classes using the available attributes. If only those attributes that preserve the indiscernibility relation are retained, any remaining attributes are redundant since their omission will not affect classification. There are usually many such subsets of attributes; however, those which are minimal are termed minimal reducts. A minimal reduct is therefore a minimal set of attributes that preserves the partitioning of the universe and hence the ability to perform the same classification as the complete data set. In practical terms, this means that no attributes can be removed from the subset without affecting the dependency measure. If $\mathbf{R}$ be the set of all reducts, then minimal reducts $\mathbf{R_{min}} \subseteq \mathbf{R}$ can be defined as:

$$\mathbf{R}_{\min} = \{X : X \in \mathbf{R}, \forall Y \in \mathbf{R}, |X| \leq |Y|\} \tag{16}$$

The search for minimal reducts is, however, non-trivial (Skowron & Rauszer, 1992; Swiniarski & Skowron, 2003), and it can be demonstrated that the number of reducts for a given information system with $n$ attributes can be as much as:

$$\binom{m}{\lfloor m/2 \rfloor} \tag{17}$$

The intersection of all the sets in $\mathbf{R}$ is termed the *core*. This set contains the attributes which cannot be eliminated without the introduction of contradictions in the data.

Many rough set approaches for dealing with data opt for search techniques which tend to balance the need for the discovery of minimal reducts with the computational overhead involved in searching for such reducts. The greedy hill-climbing search (Chouchoulas & Shen, 2001) is such

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

an example, and although it will not guarantee minimality, it is relatively efficient in terms of time/space complexity—$(n^2 + n)/2$ for a data dimensionality of $n$. Other search techniques which also do not guarantee minimality but which have been employed for the rough set methodology include backward elimination (similar to hill climbing; Dash & Liu, 1997), compound selection (Molina *et al.*, 2002), and stochastic selection (Brassard & Bratley, 1996). However, where the discovery of minimal reducts is necessary, this approach may not be acceptable, and this has frustrated efforts to apply the rough set methodology to application domains which involve large numbers of features and relatively few objects (Komorowski *et al.*, 1999) such as gene expression data.

There are various search techniques and heuristics, however, which can be used to alleviate this problem. GAs are an obvious candidate for this type of problem, and indeed the works in Jensen and Shen (2004a) and Wróblewski (1995) employ such techniques to search for minimal reducts. Although such techniques cannot guarantee minimality, they do offer an alternative which will avoid local minima. Problems may arise when employing GAs for situations where the number of data attributes is high, as the amount of time taken to discover reducts may increase considerably.

Another approach similar to GAs is particle-swarm-optimization (PSO; Wang *et al.*, 2007), which does not require operations such as crossover and mutation, but primitive and simple mathematical operators, and is also efficient in terms of time/space complexity. Again, PSO will not guarantee minimality of any reducts discovered but, like GAs, allows the search to escape local minima. Other techniques similar to GA and PSO include ACO (ant-colony-optimization; Jensen & Shen, 2004a; Jensen & Shen, 2005; Ke *et al.*, 2008) and simulated annealing (Jensen & Shen, 2004a). The approach in (Zhong *et al.*, 2001) also offers an interesting insight into the possible heuristics for finding minimal reducts.

The only way in which to ensure minimality is to conduct a complete search of all possible reducts. An exhaustive search is an example of a complete search, but it does not necessarily follow that a complete search must be exhaustive. A branch-and-bound search (Narendra & Fukunaga, 1977) is typical of a complete search that is non-exhaustive, whereas others include Boolean propositional satisfiability (SAT; Davis *et al.*, 1962). In Jensen and Shen (2008), the authors use a SAT solver algorithm (Davis *et al.*, 1962) to perform a complete search for rough set reducts. The SAT algorithm can be used to perform a complete search of the feature space and thus discover minimal reducts. Although the SAT problem is *NP – complete*, in practice the technique is both computationally efficient and can guarantee the minimality of any discovered reduct. One of the principal drawbacks of SAT, however, is that it can only be applied to discrete data domains.

## 3 Rough set extensions

The simplicity of the rough set approach is undoubtedly one of the main reasons for its success. The two areas which are most often exploited in order to extend the approach are the equivalence relation, and the subset operator, and these aspects are therefore the subject of a number of extensions. In addition to these extensions, there is also a third aspect of RST which has been exploited, that of the use of the information contained in the boundary region, or region of uncertainty. The illustration in Figure 1 shows the main RST extensions in relation to the aspects of the theory they extend to. The approaches are discussed here with reference to their underlying concepts as well as their respective merits and drawbacks.

### *3.1 Variable precision rough sets*

The VPRS approach (Ziarko, 1993) extends RST by relaxing the subset operator. It was originally proposed in order to analyse and identify data patterns which represent statistical trends rather than those which are functional. At the heart of VPRS is the idea of allowing objects to be classified with an error smaller than a given predefined level or threshold. The introduction of this threshold means that, unlike the traditional rough set approach, VPRS requires additional information other than that contained within the data.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

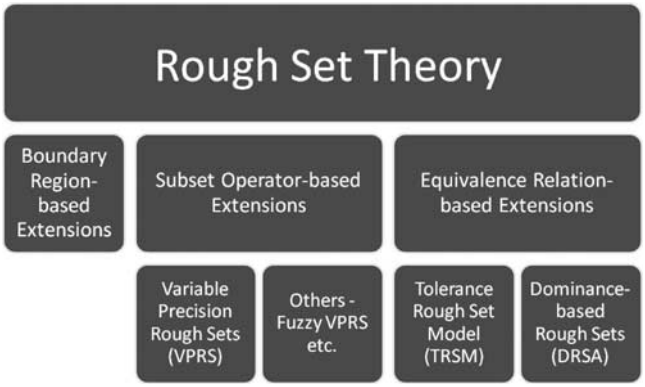



**Figure 1** A taxonomy of rough set extensions

If $X, Y \subseteq \mathbb{U}$, then the relative classification error is defined by:

$$c(X, Y) = 1 - \frac{|X \cap Y|}{|X|} \tag{18}$$

Note that $c(X,Y) = 0$ if and only if $X \subseteq Y$. A degree of inclusion can therefore be achieved by allowing a certain level of error, $\beta$, in classification:

$$X \subseteq_\beta Y \Leftrightarrow c(X, Y) \leq \beta,\ 0 \leq \beta \leq 0.5 \tag{19}$$

Thus, by replacing $\subseteq$ with the operator $\subseteq_\beta$, the $\beta$-upper and $\beta$-lower approximations can be formulated:

$$\underline{R}_\beta X = \{x|\ \ [x]_R \subseteq_\beta X\} \tag{20}$$

$$\overline{R}_\beta X = \{x|\ \ c([x]_R, X) < 1 - \beta\} \tag{21}$$

Note that when $\beta = 0$, $\underline{R}_\beta X = \underline{R}X$.

Using this extension, the positive, negative, and boundary regions can now also be defined:

$$POS_{R\beta}(Q) = \bigcup_{X \in U/Q} \underline{R}_\beta X \tag{22}$$

$$NEG_{R\beta}(Q) = U - \bigcup_{X \in Q} \overline{R}_\beta X \tag{23}$$

$$BND_{R\beta}(Q) = \bigcup_{X \in Q} \overline{R}_\beta X - \bigcup_{X \in Q} \underline{R}_\beta X \tag{24}$$

Returning to the example data set in Table 1, Equation (22) can be used to calculate the $\beta$-positive region for $R = \{b, c\}$, $X = \{e\}$, and $\beta = 0.4$. Setting $\beta$ to this value means that a set is considered to be a subset of another if they share about half the number of elements. The partitions of the universe of objects for $R$ and $X$ are:

$$\mathbb{U}/R = \{\{1, 4\}, \{2, 6, 7\}, \{3\}, \{5\}\}$$
$$\mathbb{U}/X = \{\{1\}, \{2, 3, 6\}, \{4, 5, 7\}\}$$

For each set $A \in \mathbb{U}/R$ and $B \in \mathbb{U}/X$, the value of $c(A, B)$ must be less than $\beta$ if the equivalence class $A$ is to be included in the $\beta$-positive region. Considering $A = 5$ gives

$$c(\{5\}, \{1\}) = 1 > \beta$$
$$c(\{5\}, \{2, 3, 6\}) = 1 > \beta$$
$$c(\{5\}, \{4, 5, 7\}) = 0 < \beta$$

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

Therefore, object 5 is added to the $\beta$-positive region as it is a $\beta$-subset of $\{4, 5, 7\}$ (and is in fact a traditional subset of the equivalence class). Taking $A = \{2, 6, 7\}$, a more interesting case is presented:

$$c(\{2, 6, 7\}, \{1\}) = 1 > \beta$$
$$c(\{2, 6, 7\}, \{2, 3, 6\}) = 0.3333 < \beta$$
$$c(\{2, 6, 7\}, \{4, 5, 7\}) = 0.6667 > \beta$$

Here, the objects 2, 6, and 7 are included in the $\beta$-positive region as the set $\{2, 6, 7\}$ is a $\beta$-subset of $\{2, 3, 6\}$. Calculating the subsets in this way leads to the following $\beta$-positive region:

$$POS_{R,\beta}(X) = \{2, 3, 5, 6, 7\}$$

Compare this with the positive region generated previously: $\{3, 5\}$. Objects 2, 6, and 7 are now included due to the relaxation of the subset operator. Consider a decision table $(\mathbb{U}, \mathbb{C} \cup \mathbb{D})$, where $\mathbb{C}$ is the set of conditional attributes and $\mathbb{D}$ the set of decision attributes. The $\beta$-positive region of an equivalence relation $Q$ on $\mathbb{U}$ may be determined by:

$$POS_{R,\beta}(Q) = \bigcup_{X \in \mathbb{U}/Q} \underline{R}_\beta X$$

A more comprehensive investigation of reducts for the VPRS approach may be found in Beynon (2000, 2001), and Kryszkiewicz (1994). No general comparative studies appear to have been carried out with regard to comparing the rough set and the VPRS methods, although in Thangavel *et al.* (2006), the authors compare feature selection methods based on both RST and VPRS.

As indicated previously, the VPRS approach requires the specification of an additional parameter ($\beta$). This parameter can be approximated by repeated experimentation. However, problems may arise if searching for true reducts, as the VPRS approach incorporates an element of inaccuracy in determining the number of classifiable objects.

### 3.2 *Tolerance rough sets*

The TRSM (Skowron & Stepaniuk, 1996) can be useful for application to real-valued data. TRSM employs a similarity relation to minimize data as opposed to the indiscernibility relation used in classical rough sets. This allows a relaxation in the way equivalence classes are considered. The effect of employing this relaxation means that the granularity of the rough equivalence classes has been blurred slightly. This flexibility enables a change to occur in the boundaries of the former rough or crisp equivalence classes and objects may now belong to more than one so-called tolerance class which is TRSM equivalent of a rough set equivalence class.

The tolerance threshold ($\tau$) is a global similarity threshold which determines the required level of similarity for inclusion within a tolerance class. The specification of this threshold, however, is a departure from the traditional rough set approach, which relies only upon the information contained in the data.

In this approach, suitable similarity relations must be defined for each feature, although the same definition can be used for all features if applicable. A standard measure for this purpose, given in Skowron and Stepaniuk (1996), is:

$$SIM_a(x, y) = 1 - \frac{|\, a(x) - a(y) \,|}{|\, a_{\max} - a_{\min} \,|} \tag{25}$$

where $a$ is a considered feature, and $a_{\max}$ and $a_{\min}$ denote the maximum and minimum values of $a$, respectively. When considering the case where there is more than one feature, the defined similarities must be combined to provide an overall measure of similarity of objects. For a subset of features, $P$, this can be achieved in many ways, including the following approaches:

$$(x, y) \in SIM_{P,\tau} \iff \prod_{a \in P} SIM_a(x, y) \geq \tau \tag{26}$$

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

**Table 2** Real-valued data—example

| Object | $a$ | $b$ | $c$ | $f$ |
|---|---|---|---|---|
| 0 | −0.4 | −0.3 | −0.5 | *No* |
| 1 | −0.4 | 0.2 | −0.1 | *Yes* |
| 2 | −0.3 | −0.4 | −0.3 | *No* |
| 3 | 0.3 | −0.3 | 0 | *Yes* |
| 4 | 0.2 | −0.3 | 0 | *Yes* |
| 5 | 0.2 | 0 | 0 | *No* |

$$(x, y) \in SIM_{P,\tau} \iff \frac{\sum_{a \in P} SIM_a(x, y)}{|P|} \geq \tau \tag{27}$$

where $\tau$ is a global similarity threshold which determines the required level of similarity for inclusion within a tolerance class. The framework also allows for the specific case of traditional rough sets by defining a suitable similarity measure (e.g. complete equality of features) and threshold ($\tau = 1$). Further similarity relations are summarized in Nguyen and Skowron (1997a), but are not included here. From this, the tolerance classes that are generated by a given similarity relation for an object $x$ are defined as:

$$SIM_{P,\tau}(x) = \{y \in U \,|\, (x, y) \in SIM_{P,\tau}\} \tag{28}$$

Lower and upper approximations are defined in a similar way to those of traditional RST:

$$\underline{P_\tau} X = \{x \,|\, SIM_{P,\tau}(x) \subseteq X\} \tag{29}$$

$$\overline{P_\tau} X = \{x \,|\, SIM_{P,\tau}(x) \cap X \neq \theta\} \tag{30}$$

The tuple $\langle \underline{P_\tau} X, \overline{P_\tau} X \rangle$ is known as a tolerance rough set (Skowron & Stepaniuk, 1994). Using this, the positive region and dependency functions can be defined as follows:

$$POS_{P,\tau}(Q) = \bigcup_{X \in U/Q} \underline{P_\tau} X \tag{31}$$

$$\gamma_{P,\tau}(Q) = \frac{|POS_{P,\tau}(Q)|}{|U|} \tag{32}$$

These definitions are analogous to the traditional rough set concepts and can be applied in the same way as demonstrated in Section 2.1.1. To demonstrate the approach, a sample data set is included in Table 2, which has three real-valued conditional attributes and a single crisp-valued decision attribute.

For this example, the similarity measure is the same as that given in 26 for all conditional attributes, with $\tau = 0.8$. The choice of this threshold allows attribute values to differ to a limited degree, with close values considered as though they are identical.
Thus, by making $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, and $F = \{f\}$, the following tolerance classes are generated:

$$\begin{aligned}
\mathbb{U}/SIM_{A,\tau} &= \{\{0, 1, 2\}, \{3, 4, 5\}\} \\
\mathbb{U}/SIM_{B,\tau} &= \{\{0, 2, 3, 4\}, \{1\}, \{5\}\} \\
\mathbb{U}/SIM_{C,\tau} &= \{\{0\}, \{1\}, \{3, 4, 5\}, \{2\}\} \\
\mathbb{U}/SIM_{F,\tau} &= \{\{0, 2, 5\}, \{1, 3, 4\}\} \\
\mathbb{U}/SIM_{\{a,b\},\tau} &= \{\{0, 2\}, \{1\}, \{3, 4\}, \{3, 4, 5\}, \{4, 5\}\} \\
\mathbb{U}/SIM_{\{a,c\},\tau} &= \{\{0\}, \{1\}, \{2\}, \{3, 4, 5\}, \{3, 4, 5\}\} \\
\mathbb{U}/SIM_{\{b,c\},\tau} &= \{\{0, 2\}, \{1\}, \{3, 4\}, \{5\}\} \\
\mathbb{U}/SIM_{\{a,b,c\},\tau} &= \{\{0\}, \{1\}, \{2\}, \{3, 4\}, \{4, 5\}\}
\end{aligned}$$

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

It is apparent that some objects belong to more than one tolerance class. This is a result of employing a similarity measure rather than the strict equivalence of the conventional rough set model. Using these partitions, a degree of dependency can be calculated for attribute subsets, providing an evaluation of their significance in the same way as previously outlined for the crisp rough case.

From the lower approximation, the positive and boundary regions can then be generated thus:

$$POS_{B,\tau}(F) = \bigcup_{X \in U/F} \underline{B_\tau} X = \{1, 5\}$$

$$BND_{B,\tau}(F) = \bigcup_{X \in U/F} \overline{B_\tau} X - POS_{B,\tau}(F) = \{0, 2, 3, 4\}$$

These concepts can be then employed in the same way as those of traditional rough sets to partition the data.

### *3.3 Dominance-based rough sets*

The dominance-based rough set approach (DRSA; Greco *et al.*, 2001) is an extension of RST for multi-criteria decision analysis. In contrast to traditional RST, DRSA employs a dominance relation instead of an equivalence relation. This allows DRSA to deal with the inconsistencies which are typical of criteria- and preference-ordered decision classes.

The ordering of data describing decision situations is naturally related to preferences of considered condition and decision attributes. Traditional RST does not have the ability to deal with ordinal data in the same way that DRSA does. This is because DRSA employs a dominance relation in place of the traditional rough set equivalence relation.

In DRSA, data are represented in decision table form. Let $S = \langle \mathbb{U}, Q, V, f \rangle$, where $\mathbb{U}$ is a non-empty set of finite objects, $Q$ is a finite set of criteria, and $V = \bigcup_{q \in Q} V_q$, where $V_q$ is the set of values that the criterion $q$ can take, and $f : \mathbb{U} \times Q \rightarrow V$ is an information function such that $f(x, q) \in V_q$ for every $(x, q) \in \mathbb{U} \times Q$. The set $Q$ consists of condition criteria $\mathbb{C}$, and the decision class $\mathbb{D}$. Note that $f(x, q)$ is the evaluation of object $x$ on criterion $q \in \mathbb{C}$, while $f(x, d)$ is the decision class assignment for that object.

In order for DRSA to operate effectively on pre-ordered data, the approach employs a 'preferencing' or 'outranking' relation. A typical example is: $\succeq_q$; $x \succeq_q y$, which means that $x$ is preferential to or 'outranks' $y$ with respect to $q$. The values that $q$ can take constitute a subset of real numbers—$\mathbb{R}$, such that $V_q \subseteq \mathbb{R}$, and the preference relation is a simple order between real numbers $\geq$ such that $x \succeq_q y \Longleftrightarrow f(x, q) \geq f(y, q)$ holds. This relation is straightforward for a simple maximization criterion, for example, an exam result—'the higher, the better'. For criteria where the opposite is true, for example, student failure-rate ('the less, the better'), the relation can be satisfied by negated values of $V_q$. If $P \subseteq \mathbb{C}$, it can be said that $x$ dominates $y$, denoted by $xD_P y$, if $x$ is 'better' than $y$ for every criterion from $P$, $x \succeq_q y$, $\forall q \in P$. For each $P \subseteq \mathbb{C}$, the dominance relation $D_P$ is reflexive and transitive. Given that $P \subseteq \mathbb{C}$ and $x \in \mathbb{U}$,

$$D_P^+(x) = \{y \in \mathbb{U} : yD_p x\} \tag{33}$$

$$D_P^-(x) = \{y \in \mathbb{U} : xD_p y\} \tag{34}$$

These are termed the $P$-dominating set and $P$-dominated set, respectively.

As the DRSA deals with ordinal data and objects, the manipulation of the data is carried out with respect to the ranking of decision classes. Let $T = \{1, \ldots, n\}$. The domain values of decision criterion, $V_d$, consist of $n$ elements (it is assumed that $V_d = T$) and induce a partition of $\mathbb{U}$ into $n$ classes $Dc = \{Dc_t, t \in T\}$, where $Dc_t = \{x \in \mathbb{U} : f(x, d) = t\}$. Each object $x \in \mathbb{U}$ is assigned to only one decision class $Dc_t, t \in T$. All of the classes are preference-ordered according to an increasing order of class indices, that is, $\forall r, s \in T \mid r \geq s$, objects from $Dc_r$ are preferential to the objects from $Dc_s$. Thus, the upward and downward unions of classes, respectively, can be defined as:

$$Dc_t^{\geq} = \bigcup_{s \geq t} Dc_s \quad Dc_t^{\leq} \bigcup_{s \leq t} Dc_s \quad t \in T \tag{35}$$

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

**Table 3** Dominance-based rough set approach—example

| Object | *a* | *b* | *c* | *f* |
|---|---|---|---|---|
| 0 | C | C | G | *q* |
| 1 | D | C | G | *r* |
| 2 | C | D | G | *r* |
| 3 | G | C | D | *q* |
| 4 | G | G | C | *q* |
| 5 | G | C | C | *r* |
| 6 | D | D | G | *s* |
| 7 | D | C | C | *r* |
| 8 | C | C | D | *s* |
| 9 | D | C | D | *s* |

In DRSA, the knowledge being approximated is a collection of upward and downward unions of decision classes. The knowledge granules employed for approximation in DRSA are the *P*-dominating and *P*-dominated sets, which are analogous to the equivalence classes of traditional RST. The *P*-lower and the *P*-upper approximation of $Dc_t^{\geq}, t \in T$ are denoted $\underline{P}(Dc_t^{\geq})$ and $\overline{P}(Dc_t^{\geq})$, respectively, and can be defined as follows:

$$\underline{P}(Dc_t^{\geq}) = \{x \in U : D_P^+(x) \subseteq Dc_t^{\geq}\} \tag{36}$$

$$\overline{P}(Dc_t^{\geq}) = \{x \in U : D_P^-(x) \cap Dc_t^{\geq} \neq \theta\} \tag{37}$$

Similarly, the *P*-lower and the *P*-upper approximation of $Dc_t^{\leq}$, denoted by $\underline{P}(Dc_t^{\leq})$ and $\overline{P}(Dc_t^{\leq})$, respectively, can be defined thus:

$$\underline{P}(Dc_t^{\leq}) = \{x \in U : D_P^-(x) \subseteq Dc_t^{\leq}\} \tag{38}$$

$$\overline{P}(Dc_t^{\leq}) = \{x \in U : D_P^+(x) \cap Dc_t^{\leq} \neq \theta\} \tag{39}$$

As with traditional RST, the boundary regions of $Dc_t^{\geq}$ and $Dc_t^{\leq}$ can also be defined:

$$BND_P(Dc_t^{\geq}) = \overline{P}(Dc_t^{\geq}) - \underline{P}(Dc_t^{\geq}) \tag{40}$$

$$BND_P(Dc_t^{\leq}) = \overline{P}(Dc_t^{\leq}) - \underline{P}(Dc_t^{\leq}) \tag{41}$$

To demonstrate the basic concepts of the dominance rough set approach, a small example is shown here. The example data set in Table 3 has three conditional attributes (a, b, c) and one decision attribute (f) according to the decision attribute, the objects are divided into three preference-ordered classes: $Cls1 = \{q\}$, $Cls2 = \{r\}$, and $Cls3 = \{s\}$. Thus, the following unions of classes can be approximated:

- $Cls_1^{\leq}$—the class of (at most) *q* objects
- $Cls_2^{\leq}$—the class of at most *r* objects
- $Cls_2^{\geq}$—the class of at least *r* objects
- $Cls_3^{\geq}$—the class of (at least) *s* objects

The lower approximations of the class unions consist of the following objects:

- $\underline{P}(Cls_1^{\leq}) = \{0, 4\}$
- $\underline{P}(Cls_2^{\leq}) = \{0, 1, 2, 3, 4, 5, 7\} = Cls_2^{\leq}$
- $\underline{P}(Cls_2^{\geq}) = \{1, 2, 6, 7, 8, 9\}$
- $\underline{P}(Cls_3^{\geq}) = \{6, 8, 9\} = Cls_3^{\geq}$

Therefore, only classes $Cls_1^{\leq}$ and $Cls_2^{\geq}$ cannot be approximated without uncertainty. The upper approximations can be shown to be:

$$\overline{P}(Cls_1^{\leq}) = \{0, 3, 4, 5\}$$

$$\overline{P}(Cls_2^{\geq}) = \{1, 2, 3, 5, 6, 7, 8, 9\}$$

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

While the boundary regions for $Cls_1^{\leq}$ and $Cls_2^{\geq}$ are:

$$BND_P(Cls_1^{\leq}) = BND_P(Cls_2^{\geq}) = \{3, 5\}$$

These concepts can be used in a similar way to those of traditional RST in order to deal with ordinal data.

### 3.4 *Vaguely quantified rough sets*

In traditional RST, an object is a member of the upper approximation of a set if it is related to one of the elements in the set, while the lower approximation only retains those objects related to all the elements in the set. This is a result of the use of an existential quantifier in the definition of the upper approximation, and the use of a universal quantifier for the lower approximation. For real world data which include noise to a greater or lesser degree, this approach will inevitably suffer from classification errors and inconsistency. The associated definition of the upper approximation may be too general (thus resulting in very large sets), while the definition of lower approximation might be too rigid (resulting in an empty set in the extreme case). Fuzzy RST (which is covered in the next section) exhibits similar behaviour where the quantifiers $\exists$ and $\forall$ are replaced by the *sup* and *inf* operations (Cornelis *et al.*, 2007). These operators, however, can be as susceptible to the effects of noise as their crisp counterparts.

As demonstrated previously in Section 3.1, thresholds are introduced in VPRS to deal with these problems for the crisp case. In general, given $0 < l < u < 1$, an element $y$ is added to the lower approximation of a set $A$ if at least (100 $x$ $u$)% of the elements related to $y$ are in $A$. Likewise, $y$ belongs to the upper approximation of $A$ if more than (100 $x$ $l$)% of the elements related to $y$. This can be interpreted as a generalization of the rough set model using crisp quantifiers *at least* (100 $x$ $u$)%, and *more than* (100 $x$ $l$)% to replace the universal quantifier which demands rigid (at least 100%) membership for an element to be included in the lower approximation, and the existential quantifier which demands membership that is non-zero (greater than 0%) for an element to be included in the upper approximation.

In perhaps what is one of the most recent extensions of rough sets, the authors of (Cornelis *et al.*, 2007) introduce vague quantifiers like 'most' and 'some' to the rough set model. As a result of this, an element $y$ now belongs to the lower approximation of $A$ if most of the elements related to $y$ are included in $A$. Similarly, an element belongs to the upper approximation of $A$ if some of the elements related to $y$ are included in $A$. In addition, the vague quantifiers are modelled mathematically in terms of the notion of fuzzy quantifiers in Zadeh (1965), so that the VQRS model inherits not only the flexibility of VPRS for dealing with classification errors mentioned previously, but also that of fuzzy sets for the expression of partial constraint satisfaction—by distinguishing between varying levels of membership of both the upper and lower approximations.

The definitions used for the upper and lower approximations in VPRS can be relaxed, through the use of vague quantifiers, to express that $y$ belongs to the upper approximation of the set $X$ to the extent that some elements of $y$'s equivalence class ($Ry$) are in the set $A$, and $y$ belongs to the lower approximation of $A$ to the extent that most elements of $Ry$ are in $X$. In VQRS, it is implicitly assumed that the approximations are fuzzy sets, that is, mapped from $X$ to [0, 1], that evaluate the degree to which the associated condition is fulfilled. The concept of a fuzzy quantifier in Zadeh (1965) is employed, that is, a [0, 1]→[0, 1] mapping $Q$. The set $Q$ is said to be regularly increasing, if it is increasing *and* it satisfies the boundary conditions $Q(0) = 0$ and $Q(1) = 1$. Examples of fuzzy quantifiers can be generated by means of the following parameterized formula, for $0 \leq \alpha < \beta \leq 1$, and $x \in [0, 1]$,

$$Q_{(\alpha,\beta)}(x) = \begin{cases} 0, & x \leq \alpha \\ \frac{2(x-\alpha)^2}{(\beta-\alpha)^2}, & \alpha \leq x \leq \frac{\alpha+\beta}{2} \\ 1 - \frac{2(x-\beta)^2}{(\beta-\alpha)^2}, & \frac{\alpha+\beta}{2} \leq x \leq \beta \\ 1, & \beta \leq x \end{cases}. \tag{42}$$

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

For instance, $Q_{(0.1,\ 0.6)}$ and $Q_{(0.2,\ 1)}$ may be used to reflect the vague quantifiers *some* and *most*, respectively, from natural language.

The VQRS upper and lower approximations can be defined once the quantifier pair $(Q_l, Q_u)$ has been fixed such that:

$$\mu_{\underline{R_P}X}^{Q_u}(y) = Q_u\left(\frac{|R_P y \cap X|}{|R_P y|}\right) \tag{43}$$

$$\mu_{\overline{R_P}X}^{Q_l}(y) = Q_l\left(\frac{|R_P y \cap X|}{|R_P y|}\right) \tag{44}$$

In other words, an element $y$ belongs to the lower approximation of $X$ if most of the elements related to $y$ are included in $X$. Likewise, an element belongs to the upper approximation of $X$ if some of the elements related to $y$ are included in $X$. Notice that when $X$ and $R_P$ are a crisp set and a crisp equivalence relation, respectively, the approximations may still be non-crisp because of the use of vague quantifiers. In the interests of brevity, and due to significant overlap with fuzzy-rough sets, an example is not included here. Further detail and examples of VQRS, however, are covered in (Cornelis & Jensen, 2008).

### 3.5 Other rough set extensions

As mentioned previously, perhaps one of the most appealing aspects of traditional RST lies in its simplicity. It is based on straightforward set operations and is computationally efficient. Examining the concepts described earlier in Section 2.1, the most obvious areas for further exploration and extension are the equivalence relation and the subset operator, both of which are extended by the VPRS/VQRS and TRSM/DRSA approaches, respectively. One possible avenue for further exploration which has not been examined previously lies in a variable precision tolerance rough set approach. Although this would involve the specification of two parameters, it could take advantage of the benefits offered by both models: the ability to deal with real-valued data from TRSM and the ability to handle noise from the VPRS approach.

There is also one further aspect of RST, however, that is often overlooked: the upper approximation concept and its potential contribution to improving the performance of the rough set model. Work in this area has included an approach which generates reducts that preserve the rough upper approximation (Inuiguchi & Tsurumi, 2006), as well as an approach that considers the upper approximation and proposes a feature selection algorithm based on a rough upper approximation measure (Deogun *et al.*, 1995).

Other techniques, such as those presented in Hu *et al.* (2007a), Mac Parthaláin *et al.* (2007), and Mac Parthaláin and Shen (2009), consider the positive and boundary regions as conceptually different entities, and attempt to use the boundary region information for both feature selection and classification.

In particular, in Hu *et al.* (2007a), the authors employ a consistency measure for feature selection in order to determine the classification of objects in the rough set boundary region and use this information to search for reducts. The approach uses a greedy-type search to select attributes which result in the greatest increase in the consistency value. Problems may arise, however, if the data on which the approach is operating are inconsistent; in these cases, a stopping threshold must be specified to avoid overfitting.

The approach in Mac Parthaláin *et al.* (2007), however, treats the data in the same way as those of traditional crisp RST. The central idea of this approach is that, from an intuitive point-of-view, objects in the boundary region of a given concept are more likely to belong to that concept if they are *close* to the objects of the positive region. Thus, a distance measure is employed to determine the '*closeness*' or proximity of boundary region objects to those objects in the positive region. This proximity information is then used in feature selection as a measure to determine the 'goodness' or value of potential reducts.

An approach which examines the boundary region of tolerance rough sets (and thus can also handle real-valued data) based on Mac Parthaláin *et al.* (2007) has also been proposed (Mac Parthaláin & Shen,

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

2009). Also, in Nguyen and Slezak (2004), the authors discuss what they term 'approximate reducts', based on exploiting the rough set boundary. However, the work does not outline their application.

Another interesting idea which is explored in Slowinski and Vanderpooten (1997) and Slowinski and Vanderpooten (2000) is the re-definition of the upper and lower approximation concepts of RST. The definitions propose the use of fuzzy similarity, and tolerance, as opposed to indiscernibility, although otherwise the framework remains unchanged from that of traditional RST. Similar treatment is also given by the authors in Zhao and Zhang (2005) to VPRS to extend the $\beta$-upper and $\beta$-lower approximations; however, only similarity is explored in this case.

## 4 Combining rough sets with other techniques

The combination of RST with other soft computing techniques to form hybrid systems has highlighted the value of employing RST as a part of a wider framework for improving the overall performance of such systems. Such hybrids include the combination of RST with neural networks, GAs, evolutionary algorithms, and fuzzy sets. Very significantly, there is the hybridization of rough sets and fuzzy sets to form fuzzy-RST.

### *4.1 Rough set hybridizations*

It has been demonstrated that RST can be very effective for preprocessing data input for neural networks (Jelonek *et al.*, 1994). More recent work (Mak & Munakata, 2002) has compared the rule extraction capabilities of both rough sets and neural networks and hybrid methods with ID3. The work of Yahia *et al.* (2000) further reinforces the utility of employing RST either as a neural network's preprocessor or as a combined inference mechanism for medical diagnosis and is tested on a hepatitis disease data set. Another approach for medical image classification is reported in Shang and Shen (2002) that uses RST as a dimensionality reduction step before the application of a neural networks based classifier. Further detail with regard to the use of rough sets and hybrid methods for medical applications can be found in Pattaraintakorn and Cercone (2007).

In Li and Wang (2004), a hybrid rough set and neural networks approach for rule induction is presented. This technique is applied to relatively large data sets in order to generate more concise and accurate rules than either neural networks or rough sets alone. A feature selection algorithm is proposed and rules are generated from a decision table based on the rough set discernibility matrix. Reducts and rules are obtained using RST with neural networks employed to remove noisy data. Other rough set/neural network hybrid approaches are also to be found in Jelonek *et al.* (1994), Mitra and Banerjee (1996), Swiniarski *et al.* (1995), and Wang *et al.* (2005). In addition, it has been demonstrated that rough sets can help to generate new models of neurons in Lingras (1996, 1997).

A review of the hybridization of RST with GAs is documented in Cordon *et al.* (2001). Prior to this, the first hybridization based on lower and upper bounds of numeric ranges was proposed as a rough-GA in Lingras and Davies (2001). Others include: genetic encoding in order to generate rough set representations of clusters Lingras and West (2004), and a hybrid decision support system for cancer detection Mitra and Mitra (2000). Genetic programming has also been allied with rough sets for bankruptcy classification McKee and Lensberg (2002).

RST has also been hybridized with classical statistical methods such as PCA (Swiniarski, 1999), Bayesian methods (Swiniarski, 1998), or wavelets (Wojdyllo, 1998). Such integration has resulted in classifiers of better quality than those constructed through the use of RST alone (Browne *et al.*, 1998).

In terms of hybridizing rough set extensions, a number of approaches have been proposed, such as fuzzy-rough VPRS (Mieszkowicz-Rolka & Rolka, 2004), and dominance-based rough sets and VPRS (Hu & Yu, 2004). An interesting idea that has not yet been explored is a VPRS and TRSM hybrid. This would allow the flexibility to deal with real-valued data inherited from the TRSM approach and the noise tolerance of the VPRS method. This would mean the specification of two parameters, however, which would involve significant experimentation in order to establish ideal values for a given set of data.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

### *4.2 Fuzzy-rough sets*

FST was first proposed nearly 44 years ago (Zadeh, 1965) and RST will celebrate its 28th anniversary this year (Pawlak, 1981). FST and RST complement one another (Dubois & Prade, 1992) and much advantage has been taken of this fact. This is reflected in the breadth and depth of research which has been undertaken in this particular hybridization of rough sets.

Note that fuzzy-rough sets should not be confused with existing approaches that directly combine the use of RST for dimensionality reduction and that of FST for knowledge modelling for example (Shen & Chouchoulas, 2002; Shan *et al.*, 2002). Although successful in real-world applications, the underlying ideas of such work are straightforward and hence are omitted from the discussions below.

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 (Yeung *et al.*, 2005). For the constructive approach, generalized lower and upper approximations are defined based on fuzzy relations. Initially, these were fuzzy similarity/equivalence relations (Dubois & Prade, 1992) but have since been extended to arbitrary fuzzy relations (Yeung *et al.*, 2005). The axiomatic approach is primarily for the study of the mathematical properties of fuzzy-rough sets (Wu & Zhang, 2004).

In (Dubois & Prade, 1992), the authors define the fuzzy $P$-lower and $P$-upper approximations as follows:

$$\mu_{\underline{P}X}(F_i) = \inf_x \max\{1 - \mu_{F_i}(x), \mu_X(x)\} \quad \forall i \tag{45}$$

$$\mu_{\overline{P}X}(F_i) = \sup_x \max\{\mu_{F_i}(x), \mu_X(x)\} \quad \forall i \tag{46}$$

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 known as a fuzzy-rough set. Also in the literature are definitions for rough-fuzzy sets (Dubois & Prade, 1990; Srinivasan *et al.*, 1998), which can be seen as 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 (Yao, 1997), it is argued that, in order to remain 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, thus ensuring that both models are complementary. In this framework, the approximation of a fuzzy set in a fuzzy approximation space is considered to be a more general model, unifying both theories. However, most researchers consider the traditional definition of fuzzy-rough sets in (Dubois & Prade, 1992) as standard. The specific use of *min* and *max* operators in the above definitions is expanded in (Radzikowska & Kerre, 2002), where a wide range of fuzzy-rough sets are constructed, with each member represented by a particular implicator and t-norm. The properties of three typical implicators are investigated. Further investigations in this area can also be found in De Cock *et al.* (2004), Thiele, (1998), Wu *et al.* (2005), Yeung *et al.* (2005).

In Boixader *et al.* (2000), Morsi and Yakout, (1998), an axiomatic approach is taken, but is restricted to fuzzy T-similarity relations (and hence fuzzy T-rough sets). The properties of generalized fuzzy-rough sets are investigated in (Wu *et al.*, 2003), and a pair of dual generalized fuzzy approximation operators are defined based on arbitrary fuzzy relations. The approach presented in (Mi & Zhang, 2004) introduces definitions for generalized fuzzy lower and upper approximation operators determined by a residual implication. Assumptions are found that allow a given fuzzy set-theoretic operator to represent a lower or upper approximation from a fuzzy relation. Different types of fuzzy relations produce different classes of fuzzy-rough set algebras.

The work in (Radzikowska & Kerre, 2004) generalizes the fuzzy-rough set concept through the use of residuated lattices. An arbitrary residuated lattice is used as a basic algebraic structure, and several classes of lattice-valued fuzzy-rough sets (a fuzzy-rough hybridization of L-fuzzy sets) and their properties are investigated. In (Chen *et al.*, 2006), a complete completely distributive (CCD) lattice is selected as the foundation for defining lower and upper approximations in an attempt to provide a unified framework for rough set generalizations. It is demonstrated that the existing fuzzy-rough sets are special cases of the approximations on a CCD lattice for $T$-similarity relations.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

The relationships between fuzzy-rough set models and fuzzy topologies on a finite universe have been investigated. The first such research was reported in (Boixader *et al.*, 2000), where it was proved that the lower and upper approximation operators were fuzzy interior and closure operators, respectively, for fuzzy $T$-similarity relations. The work carried out in (Yeung *et al.*, 2005) investigated this for arbitrary fuzzy relations. In (Qina & Pei, 2005) and (Wu, 2005), it was shown that a pair of dual fuzzy rough approximation operators can induce a topological space if and only if the fuzzy relation is reflexive and transitive. The fuzzy interior (closure) operator is also examined.

In addition to the previous approaches to fuzzy-rough hybridization, other generalizations are possible. One of the first attempts at hybridizing the two theories is reported in (Wygralak, 1989), where rough sets are expressed by a fuzzy membership function to represent the negative, boundary, and positive regions. All objects in the positive region have a membership of one and those belonging to the boundary region have a membership of 0.5, while those of the negative region have a membership of 0 as they do not belong to the set of interest. Thus, in adopting this approach, a rough set can be defined using FST. This also means that the rough set operators of union and intersection are modified accordingly. In (Pedrycz, 1999), the author attempts to address the problem where the fuzzy set representation of a rough set may be too precise, such that a concept is described exactly once its membership function has been defined. The solution to this is to employ an approximation of a family of fuzzy sets which the author terms a *shadowed set*. Shadowed sets do not use exact membership values but instead use truth values and a zone of uncertainty. A similar approach to that of (Wygralak, 1989) is applied where elements may belong to a set with certainty (membership value 1), possibility (unit interval), or not belong (membership value 0). These ideas of course correspond to the rough set positive, boundary, and negative regions, respectively.

Another approach is reported in (Chimphlee *et al.*, 2006a) where the rough set lower approximation is employed, and elements are allowed to belong to this with certainty; however, the boundary region or uncertain region is fuzzified and membership values of elements are expressed in terms of a fuzzy membership function. The authors of (Mieszkowicz-Rolka & Rolka, 2004) apply a fuzzy-rough sets extension to the VPRS model described in Section 3.1 in an attempt to capitalize on the advantages of both rough sets and fuzzy sets within the VPRS framework. However, the VQRS approach of (Cornelis *et al.*, 2007) as detailed in Section 3.4 also takes advantage of these in a single approach as it employs fuzzy quantifiers and extends the VPRS approach simultaneously.

## 5 Applications

In this section, a number of theoretical and real world application areas of RST, rough set extensions, and fuzzy-RST are examined. The sheer number of applications and amount of work that has been published in the area means that it would be impossible to cover all areas in sufficient depth. Therefore, in this paper, three important areas of machine learning have been chosen for close examination; classification, clustering, and feature selection. A review of each of these areas is documented in the following sections. In each section, a further subsection is devoted to an example of real-world application.

### *5.1 Classification*

Classification concerns any problem in which a decision is taken or a forecast is made on the basis of available knowledge or information. A classification algorithm allows repeated forecasts to be made with regard to accumulated knowledge for new situations. Such algorithms can then be applied in order to classify previously unseen objects. Each new object can be assigned to a predefined set of classes, based on the observed values of suitably chosen attributes or features.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

It is interesting to note that, despite the level of interest in rough set classification which is borne out by the number of publications in the area, no comprehensive survey of rough classification has been published to date. Perhaps this is due in part to the fact that RST is often married with other approaches when applied to the classification problem. Nevertheless, a number of RST-based classifiers have been proposed. The first application of RST to the classification problem is demonstrated in (Pawlak, 1984). The authors (Pawlak & Skowron, 1993; Skowron, 1993; Slowinski *et al*., 2002) discuss the fundamentals of rough set rule induction for classification, but no algorithms are proposed.

The earliest RST-based classification algorithm is described in (Pawlak *et al*., 1986). Later examples were proposed in (Bell & Guan, 1998) and (Deogun *et al*., 1994), although the latter focused on database mining. Much use has been made of rough classifiers which were integrated into the learning from examples based on rough sets (LERS) framework (Grzymala-Busse & Grzymala-Busse, 1995; Grzymala-Busse & Wang, 1996). In these methods, descriptions of concepts are constructed through the calculation of all reducts for a given data set, by means of the decision rules. In (Bazan *et al*., 2000), it is argued that these methods are not appropriate for classifying unseen data, and thus a number of rough set classification methods are proposed which address this problem. In addition, some new methods for rule induction from reducts, as well as ways of dealing with real-valued data discretization, are also described (also within the LERS framework). Similar aspects are also examined in Grzymala-Busse (2003) and Grzymala-Busse (2006). Other research such as (Stefanowski, 1998) also concentrates on addressing some of the shortcomings of the use of rough sets for rule induction as an aid to classification.

Rough set extensions have also been employed for classification. In (Ziarko, 2003), the author discusses the use of VPRS for building decision tables from data models. Others which also employ VPRS include (Glymin & Ziarko, 2007) and (Zhao & Zhu, 2006) for email spam filtering, and general classification (Zhao *et al*., 2003). In (Wang *et al*., 2004), the authors have combined VPRS with fuzzy clustering techniques to discover rules in process planning. In the same way that VPRS has been applied to the classification task, so too has the TRSM, and a number of papers have been published in this area. Applications include handwriting classification (Kim & Bang, 2000), web document classification (Yi *et al*., 2005), and geographical land classification (Yun & Ma, 2006). Although a relatively new approach, VQRS has also been applied to the classification of mammographic data (see Section 5 for further detail; Mac Parthaláin *et al*., 2010). The DRSA has also been employed for rule induction (Shao & Zhang, 2004) and classification (Kotłowski *et al*., 2008), albeit with application to ordinal data.

Initial attempts to use fuzzy-rough sets for classification were presented in (Sarkar, 2000), which adopted a nearest neighbour (NN) type classifier approach. This approach attempted to handle both the fuzzy uncertainty due to overlapping classes and the rough uncertainty caused by a lack of informative features. A fuzzy-rough ownership function (a value which is influenced by all training objects) was employed in an effort to capture both of the aforementioned aspects. In addition, this also allows a possibilistic class membership assignment. The ownership function is influenced by all of the objects in the training set, which in turn means that the number of neighbours does not need to be defined. Other parameters must however be specified for successful operation. In (Wang *et al*., 2005a), the authors extend the approach but divide the task of classification into four parts. First, using a leave-one-out type of strategy, the fuzzy-rough ownership value is calculated for each training object for all classes. The ownership value indicates the degree to which other objects support each individual object. Inconsistencies are then filtered from the training data—a high fuzzy-rough ownership value indicates a class other than a known class. Following this, representative points are selected from the processed training data and fuzzy-rough ownership values are refreshed based on mountain clustering. Then, finally, test objects are classified using only the representative training data from the previous step using the algorithm proposed in (Sarkar, 2000).

Other NN classification methods which employ fuzzy-rough hybridization include (Bian & Mazlack, 2003), which integrates rough uncertainty into the fuzzy *k*NN classifier using the definitions

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

of fuzzy upper and lower approximations as defined in (Dubois & Prade, 1992). The membership of a test object to the upper and lower approximations for every class is determined by $k$ NN. In addition, a similar approach is used in (Mac Parthaláin *et al.*, 2010); once again, the fuzzy-rough upper and lower approximations are used to determine the membership of test objects to a particular class.

Little research has taken place in the area of fuzzy-rough decision tree induction, although there is much interest in fuzzy decision trees because of their ability to model vagueness. The work on fuzzy-rough decision trees outlined in (Bhatt & Gopal, 2004) employs the fuzzy-rough ownership measure from (Sarkar, 2000), which is used to define a 'fuzzy-roughness' measure and fuzzy-rough entropy measure. The node-splitting criterion is determined using the fuzzy-rough entropy measure. In (Jensen & Shen, 2008), a fuzzy decision tree algorithm based on the well-known fuzzy ID3 (Baldwin *et al.*, 1997) approach is described. In this case, fuzzy-rough dependency is employed to decide when node splitting should occur. An approach for rule induction using fuzzy rough sets is proposed in (Hong *et al.*, 2006) for generating certain and possible rulesets from hierarchical data.

#### *5.1.1 Image data analysis for mammographic risk assessment*

Breast cancer is a major health issue, and the most common among women in the European Union (EU). It is estimated that 8–13% of all women will develop breast cancer at some point during their lives. Furthermore, in the EU and United States, breast cancer is attributed as the leading cause of death of women in their 40s. Mammography is a process whereby low-dosage X-rays are used to generate images which can then be employed to examine the internal structure of the human breast for both diagnosis and screening. In addition to mammographic imaging, other imaging techniques such as magnetic resonance imaging (MRI) and ultrasound imaging may also be used. Although increased incidence of breast cancer has been recorded, so too has the level of early detection through screening in order to assess the risk of developing cancer using mammographic imaging and expert opinion. However, even expert radiologists can sometimes fail to detect a significant proportion of mammographic abnormalities. In addition, a large number of detected abnormalities are usually discovered to be benign following medical investigation. Existing mammographic computer-aided diagnosis (CAD) systems concentrate on the detection and classification of mammographic abnormalities. As breast tissue density increases however, the effectiveness of such systems in detecting mammographic abnormalities is reduced significantly. In addition, it is known that there is a strong correlation between mammographic breast tissue density and the risk of development of breast cancer. Automatic classification, which has the ability to consider tissue density when searching for mammographic abnormalities, is therefore highly desirable.

The approach in (Mac Parthaláin *et al.*, 2010) describes the application of a number of rough and fuzzy-rough approaches for dealing with mammographic risk assessment data. The objective of this analysis is to determine the risk of developing cancer by classifying each woman or mammogram according to a consensus class which has been agreed upon by three expert radiologists. The actual approach employs a fuzzy-rough framework. There are three steps: feature extraction to extract the features from the raw image data, feature selection which removes noisy irrelevant or redundant features from those extracted features, and classification to classify the mammograms into one of four predefined classes. The work here focuses on a brief review of the fuzzy-rough sets based classification step.

Efficient and, in particular, accurate classification of mammographic imaging is of high importance. Any improvement in accuracy for automatic mammographic classification systems can result in a reduction in the amount of required expert analysis, thus improving the time taken to perform breast abnormality risk assessment. In addition, by reducing inter-expert variation, the resulting automatic risk assessments can be more accurate. The data in mammographic imaging is real-valued and can also be noisy. Clearly, any classifier employed must therefore have the ability to deal with such data. Discrete methods require that the real-valued data are discretized and thus may result in significant information loss; however, the methods described here require no discretization, and are based on fuzzy-RST which uses only the information contained within the data.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

(1) $N \leftarrow \mathbb{U}$
(2) $\mu_1(y) \leftarrow 0, \mu_2(y) \leftarrow 0, Class \leftarrow \emptyset$
(3) $\forall X \in C$
(4) $\quad \mu_{\underline{R_P}X}(y) = \inf_{z \in N} I(\mu_{R_P}(y,z), \mu_X(z))$
(5) $\quad \mu_{\overline{R_P}X}(y) = \sup_{z \in N} T(\mu_{R_P}(y,z), \mu_X(z))$
(6) $\quad$ **if** $(\mu_{\underline{R_P}X}(y) \geq \mu_1(y)$ && $\mu_{\overline{R_P}X}(y) \geq \mu_2(y))$
(7) $\quad\quad Class \leftarrow X$
(8) $\quad\quad \mu_1(y) \leftarrow \mu_{\underline{R_P}X}(y), \mu_2(y) \leftarrow \mu_{\overline{R_P}X}(y)$
(9) **output** *Class*

**Figure 2** The FRNN algorithm (FRNN($\mathbb{U}$, $C$, $y$,): $\mathbb{U}$, the training data; $C$, the set of decision classes; $y$, the object to be classified)

The fuzzy-rough classifier employed in (Mac Parthaláin *et al.*, 2010) is based on the NN classifier technique (Jensen & Cornelis, 2008) and can be seen in Figure 2. It works on the basic principle that the lower and the upper approximations of a decision class, calculated by means of the NNs of a test object $y$, provide good clues in order to predict the membership of the test object to that class. The membership of a test object $y$ to each (crisp or fuzzy) decision class is determined via the calculation of the fuzzy lower and upper approximation. The algorithm outputs the decision class with the resulting best fuzzy lower and upper approximation memberships. The complexity of the algorithm is $O(|C| \cdot (2|\mathbb{U}|))$. Note that, although a value for the parameter $k$ that is employed in the traditional $k$NN method is not required, it can be incorporated into the algorithm to facilitate more detailed comparison by replacing by replacing line (2) with '$N \leftarrow getNearestNeighbours\ (y, k)$'.

The algorithm is applied to two mammographic imaging data sets, which have been labelled with the consensus opinion of 3 expert radiologists. The first of these is the Mammographic Image Analysis Society (MIAS) database (Suckling *et al.*, 1994), and the second is the Digital Database of Screening Mammography (DDSM; Heath *et al.*, 2000). The MIAS data set is composed of Medio-Lateral-Oblique (MLO) left and right mammograms from 161 women (322 objects). Each mammogram object is represented by 281 features extracted using the process detailed in (Oliver *et al.*, 2008). The spatial resolution of the images is 50 μm × 50 μm and is quantized to 8 bits with a linear optical density in the range 0–3.2.

The DDSM database provides four mammograms, comprising left and right MLO and left and right Cranio-Caudal (CC) views, for most women. To avoid bias, only the right MLO mammogram for each woman is selected. The data set contains 832 mammograms (objects) and again 281 features obtained in the same manner as those for the MIAS data set above.

The class labels for each mammogram are the consensus opinion of three expert radiologists. The four discrete labels ranging from 1–4, which are shown in Figure 3, relate to the BIRADS classification (American College of Radiology, 1998), where 1 represents a breast that is entirely fatty and 4 represents a breast that is extremely dense. The FRNN algorithm was compared against several other algorithms including a fuzzy NN (Keller *et al.*, 1985), a fuzzy-rough NN FRNN-O (Sarkar, 2007; based on the measure in (Sarkar, 2000)), and an approach based on VQRS (Cornelis *et al.*, 2007)—VQNN vaguely quantified the NN. The classification accuracies are obtained using 10 × 10-fold cross validation. The FRNN approach performs well compared with the other classifiers achieving accuracies of 91.2% compared with 75.12% for FNN, 82.1% for FRNN-O, and 72% for VQNN for the first data set. Values for the second data set also show that FRNN performed better than did all of the other approaches (Mac Parthaláin *et al.*, 2010).

### *5.2 Clustering*

The clustering task is the unsupervised classification of data objects (patterns observations, data vectors) into groups or clusters. Clustering has been addressed in many contexts and by researchers of many different disciplines, and this reflects its applicability and popularity as an

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

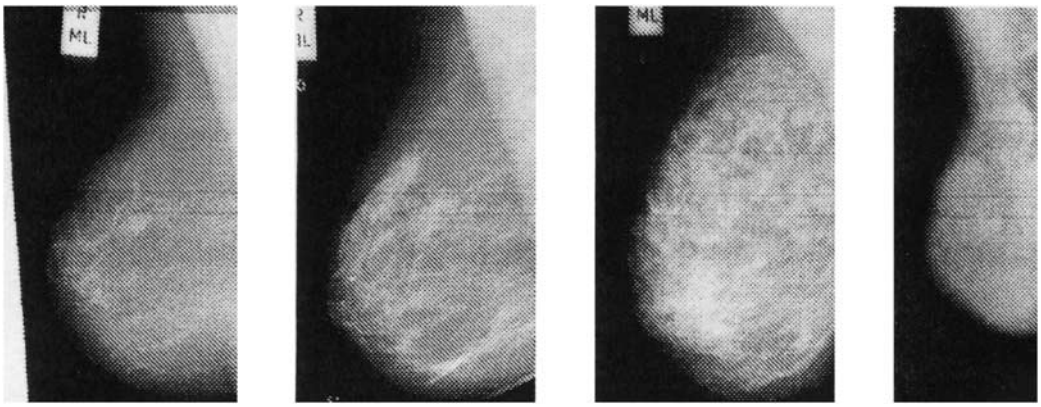


**Figure 3** Example mammograms where breast tissue density increases from L-R corresponding to BIRADS class I (far left) to class IV (far right)

important step in data analysis. Since both cluster analysis and RST form data groups, it is easy to see the conceptual similarity between the upper and lower approximation constructs of rough sets and formation of data clusters or groups. This similarity has meant that the rough sets lend themselves easily to the clustering problem. A further advantage that RST offers is that it may also provide scope for the discovery of 'possible' data clusterings through the use of the information contained in the rough set boundary region.

Much of the interest in rough clustering has been relatively recent (Hirano & Tsumoto, 2000; Hirano & Tsumoto, 2003; Peters *et al*., 2002). The application of rough sets to clustering is not limited to the use of rough indiscernibility (Hirano & Tsumoto, 2003). For instance, a rough set version of the classical *k*-means algorithm is proposed in (Lingras & West, 2004). Similarly, in (Lingras *et al*., 2004), Kohonen SOM (self-organizing maps) were used to generate intervals of clusters based on RST. The authors of (Malyszko & Stepaniuk, 2008) propose a rough set clustering algorithm by combining entropy-based thresholding with rough sets.

The use of VPRS within the framework of the fuzzy c-means (FCM) algorithm (Bezdek, 1981; Dunn, 1973) is documented in (Bao *et al*., 2006) where VPRS is employed to assign weights to each of the features. The basis for the approach is VPRS but an extension is proposed for the variable precision fuzzy-rough case. This is demonstrated by applying it to image analysis. VPRS is also used along with fuzzy-rough sets in (Zheng & Wang, 2008) as part of a fault diagnosis system. As an aid to fuzzy clustering in the general case in (Wang *et al*., 2005b), VPRS is employed for generating rules from the fuzzy conditional and decision constructs of the fuzzy clustering algorithm. Although not as popular as traditional RST or VPRS, TRSM has been applied to the clustering problem in (Kawasaki *et al*., 2000) and (Ho & Nguyen, 2002), where the authors employ an algorithm to cluster documents. Later work (Ngo & Nguyen, 2004) also used TRSM in a similar manner for clustering web search results. The traditional rough set approach is extended in (Kumar *et al*., 2007) by using a tolerance relation to form initial clusters; subsequent clusters are then formed using a constrained similarity relation which is also used as a merging criterion to combine initially identified clusters.

There have been few applications of fuzzy-RST to clustering. Most approaches, such as those of Wang *et al*. (2005b), mentioned previously, and Zhao *et al*. (2005), have tended to use both FST and RST but in isolation rather than in terms of fuzzy-RST. Rough-fuzzy sets are employed in Petrosino and Ceccarelli (2000) for texture separation in imaging, and in Pal (2004) the author also describes the application of rough-fuzzy sets for clustering and employs an image segmentation example to demonstrate this. In Chimphlee *et al*. (2006a), the authors propose a fuzzy-rough extension of the well-known FCM clustering algorithm and apply it to network security intrusion detection. Another fuzzy-rough approach which is also based on FCM is proposed in Hu and Yu (2005). It remains to be seen whether further fuzzy-rough approaches for clustering will be proposed, although it would seem that fuzzy-rough sets are well suited for such problems.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

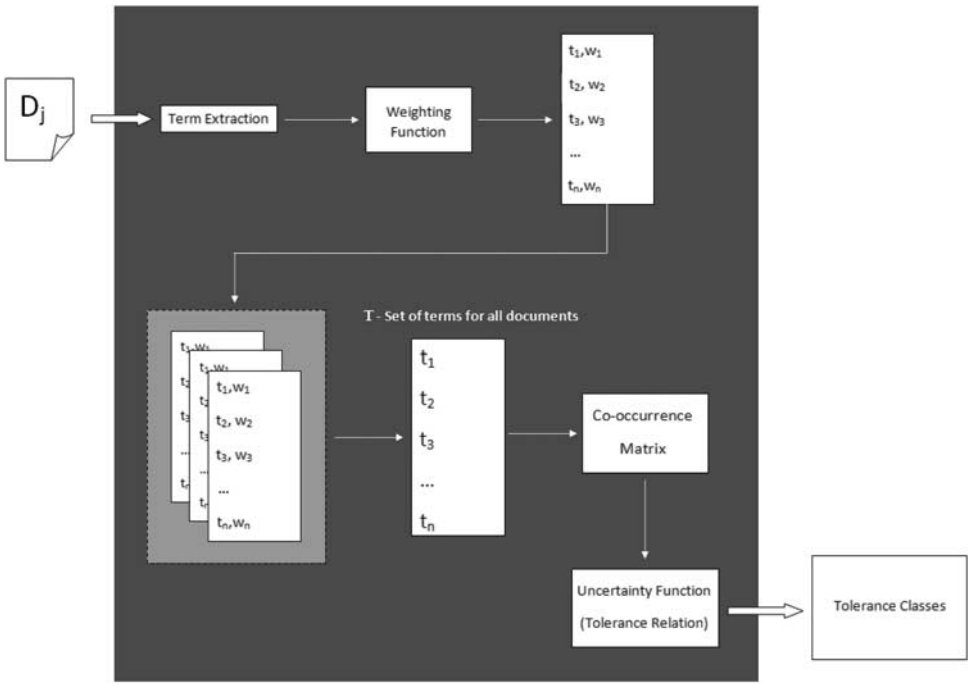



**Figure 4** Document clustering using tolerance rough sets—stage 1

*5.2.1 Document clustering*

The clustering of documents is a difficult task for a number of reasons, mainly due to the textual characteristics and unstructured format that every individual document takes. In Ho *et al.* (2006), the authors describe a method to cluster documents using tolerance rough sets. Two algorithms are described: one for hierarchical clustering and another for non-hierarchical clustering.

The approach can be broken down into two stages: the generation of tolerance classes, and the manipulation and generation of the clusters. In the first step shown below in Figure 4, a set of terms (words) is extracted from each document, and these are then assigned weights according to occurrence. Each individual term ($t_i$) is assigned a weight ($w_i$) which reflects its importance in the document; where $i = 1, 2, 3, \ldots, n$ with $n$ being the number of extracted terms. A document is denoted by $d_j = (t_1, w_{1j}; t_2, w_{2j} \ldots; t_n, w_{nj})$ and $wiq \in [0, 1]$. The weights are calculated by means of a frequency function, such that terms that occur often have a lower weight than those that rarely occur. This ensures that terms that occur in all documents have a zero weight. Each document is represented by a predefined number ($R$) of its highest weighted terms. All of the terms for all documents denoted by $T$ are used in a co-occurrence matrix to determine how terms are related to one another. Using an uncertainty function derived from a tolerance relation, this matrix can then be used to generate tolerance classes of terms in $T$. It is at this point that the tolerance value ($\tau$) must be specified for the uncertainty function.

In the second stage of the approach shown in Figure 5, a concept is defined which is used for the representation of clusters. This representation is what the authors term *polythetic* and must fulfil three properties which relate to the documents under consideration and the terms (words) in each document. Membership of each document to a cluster is defined in terms of a Bayesian minimum error rate and can be used to build each of the clusters. Cluster similarity is carried out in the usual manner, by employing a distance metric. It should be noted that clusters are built using only the upper approximation of the tolerance rough set calculated from a subset of terms $X \subseteq T$.

A number of experiments are conducted using both hierarchical and non-hierarchical clustering algorithms for both general clustering and information retrieval. In particular, the TRSM-based approach is compared with a vector space model (VSM) approach to clustering for information retrieval. This is an algebraic model for representing text documents as vectors of identifiers such as

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

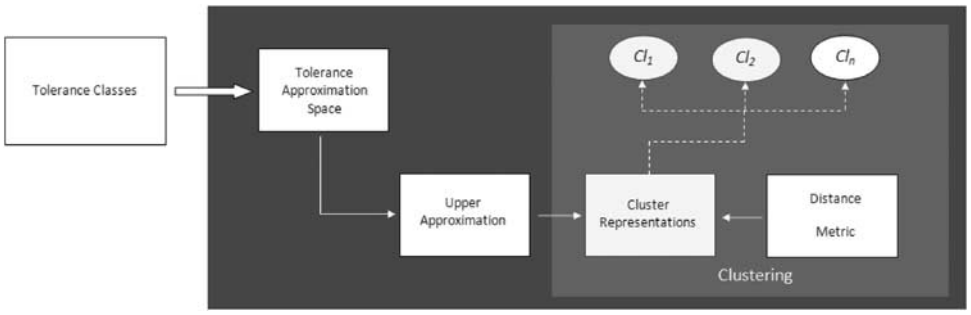



**Figure 5** Document clustering using tolerance rough sets—stage 2

index terms. The TRSM-based method demonstrates that it can equal or outperform the VSM method. This, however, requires that a range of tolerance values are specified for the uncertainty function.

It is interesting to note that there are a number of areas of this approach that could be covered by using fuzzy-RST, thus eliminating the need for the subjective specification of the thresholding value of the TRSM. In addition, a number of other thresholds relating to the number of the terms, $R$, which should be considered for each document, could also be eliminated.

### *5.3 Feature selection*

Feature selection (FS), which may also be referred to as attribute selection or semantics-preserving attribute reduction, is a term used to describe the problem of selecting input attributes that are most predictive of a given outcome. The FS problem is pervasive and can be encountered in many areas of machine learning, pattern recognition, and signal processing. In contrast to other methods for reduction of dimensionality, the FS approach preserves the original semantics or meaning of the features following reduction. FS has been applied to tasks that involve data sets which contain very large numbers of features (in the order of tens of thousands; Chouchoulas & Shen, 2001). Without FS, such problems would prove to be computationally intractable.

As RST was originally proposed for supervised learning, it is no surprise, therefore, that one of the many successful applications of RST has been in the area of FS. The basic tenet of RST, which means that only the supplied data are employed for data reduction (with no additional information), has many benefits in FS. Most other methods require at least some supplementary knowledge. The main disadvantage of rough set-based FS in the literature is the restrictive requirement for all data to be crisp, and hence the motivation to extend the rough set model as described in Section 3.

There are two main approaches when searching for rough set reducts: the dependency degree approach and the discernibility matrix approach. Both approaches have been employed for rough set-based FS, although the discernibility matrix approach is computationally expensive for large data sets (Jensen & Shen, 2008), but some constructs (Pawlak, 1991) have been proposed to alleviate this problem.

Among the earliest rough set-based dependency degree approaches to FS is the Preset algorithm (Modrzejewski, 1993), which uses RST to rank features heuristically, within the assumption of a noise-free binary domain. In Zhong *et al.* (2001), a rough set heuristic filter-based approach is presented. The algorithm starts out by calculating the core of the data set (attributes that cannot be removed without introducing inconsistency) and then it incrementally adds attributes based on a heuristic measure. A threshold value is required as a stopping criterion to determine when a reduct candidate is sufficiently 'close' to being a reduct. In Chouchoulas and Shen (2001), the authors also present a filter-based method called rough set attribute reduction (RSAR), based on the rough set dependency degree. It uses a greedy forward selection technique (starting with an empty subset) that incrementally adds features that result in an increase in the dependency value. Other approaches have also utilized this approach but used other measures such as entropy

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

(Jensen & Shen, 2004b) and a boundary region measure (Mac Parthaláin *et al*., 2007) to search for reducts. In terms of the discernibility matrix approach (Skowron & Rauszer, 1992), a number of techniques have also been proposed, and algorithms such as that described in Nguyen and Skowron (1997a) adopt this technique to search for reducts. Others also include (Øhrn, 1999) with specific application to medical problem domains, and (Wang & Wang, 2001) which attempts to address the computational complexity associated with discernibility matrices.

Although not as popular as the traditional rough set approach, VPRS has also been applied to the FS problem. In Thangavel *et al*. (2006), the authors compare VPRS and traditional rough set-based FS techniques. A fault-detection process which uses VPRS as an FS step is also described in Li *et al*. (2006). The main disadvantage with approaches like VPRS is the specification of additional tunable parameters, in this case $\beta$. As mentioned previously, the optimum value can be obtained by repeated experimentation, but this may take considerable time depending on the nature of the data being examined.

Applying rough set-based FS to domains where the data are real-valued has previously meant that the data must be discretized beforehand. Tolerance rough sets have provided a solution to this problem, however, and in (Jensen & Shen, 2008) the authors demonstrate how this can be achieved. Unfortunately, the tolerance rough set approach requires a thresholding value which is specified by the user and can only be automatically approximated by repeated experimentation. Human specification of such a threshold, however, conflicts with the rough set ideology that only the information in the data should be employed. As mentioned previously, this has resulted in the development of techniques which extend the rough set concepts of the positive region and dependency function through the use of fuzzy sets resulting in a number of fuzzy-rough set approaches (Shen & Jensen, 2004; Jensen & Shen, 2004a, 2004b, 2007, 2008, 2009; Hu *et al*., 2007b; Tsang *et al*., 2008). A greedy hill-climbing search mechanism is then employed to search for subsets of features and a new *fuzzy dependency* measure is employed as a stopping criterion. In Hu *et al*. (2006), an approach that employs information measures for fuzzy indiscernibility relations is presented for the computation of feature importance. Reducts are then calculated by employing a greedy selection algorithm. Comprehensive coverage is given to fuzzy-rough FS approaches in Jensen and Shen (2008), which explores all aspects of generation of reducts, and selection and search methods.

### *5.3.1 FS for gene expression data*

The application of techniques such as machine learning, data mining, and pattern recognition to areas of Bioinformatics has enjoyed much attention in recent years, and rough sets and their extensions are no exception. One particular area within this field is the manipulation of gene expression data. Owing to the very large number of genes in the sample data, the search space is exponentially large, and thus any techniques which are applied to this type of data must be robust. Rough set techniques are therefore an ideal candidate for the examination of such data.

Rough set FS is employed in Momin *et al*. (2006) as a dimensionality reduction step and applied to a number of gene expression data sets. The FS step generates a number of reducts which are then used to reduce the data before they are classified using a NN approach. The approach can be described as a series of individual steps as shown in Figure 6. The first step involves discretizing the data such that it can be used with the rough set approach. This discretization step involves the search for partitions for each attribute domain. These partitions form new intervals to which objects can be assigned. A Bayesian equal-width approach is used in this case, which handles outliers in a sensible manner, but assumes uniform distribution of the data.

Having discretized the data, the FS step is then implemented, using a heuristic search described below. The approach starts out with an empty set, to which those attributes that have a rough set dependency ($\gamma > 0$) are added incrementally. This generates a set of attributes from which reducts can later be generated. A thresholding value is also specified at this point termed $\lambda$; this value is used to limit the cardinality of all generated reducts. All possible reducts of cardinality $\lambda$ are then generated, but only those of $\gamma = 1$ are retained. A pruning of all super sets of reducts is then carried out, and the data are reduced prior to the next step.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

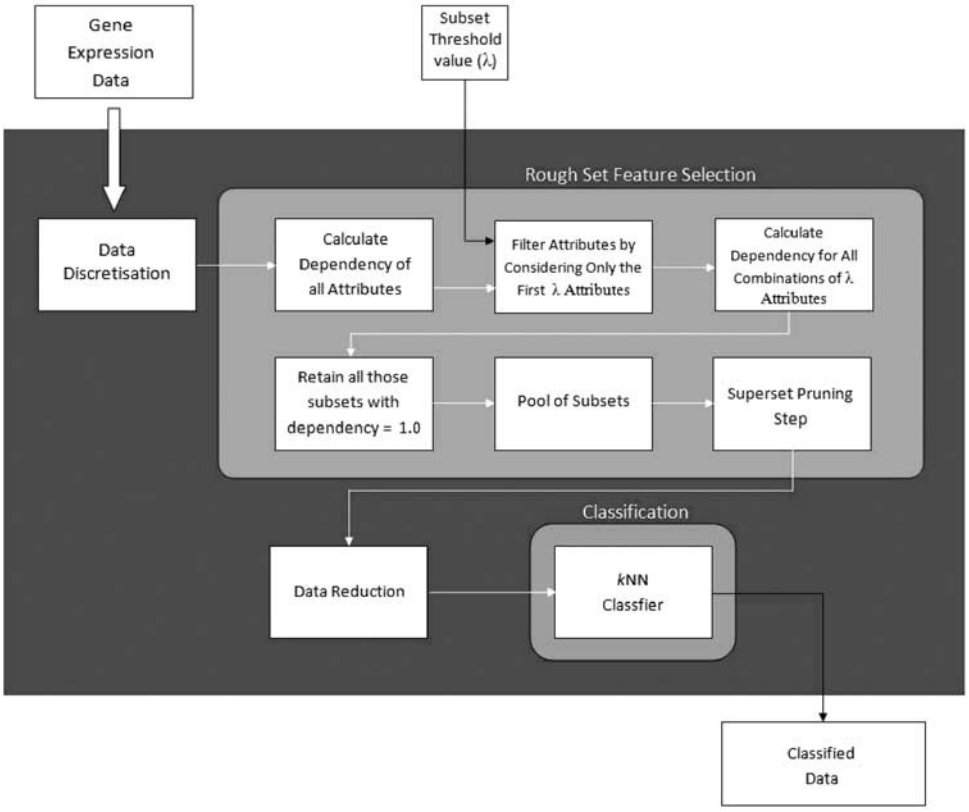



**Figure 6** Feature Selection for gene expression data

The next stage is data reduction where all of the reducts are used to minimize the data by selecting each of the features that appear in a given reduct from the data. Each reduced data set is then classified. The classifier used here is *k*NN (Keller *et al.*, 1985), which is an object-based classifier learner.

The above process is applied to four publicly available data sets relating to various types of cancer. Various values of $\lambda$ are used to generate the reducts for each of the data sets which are then classified. For the *k*NN classifier, 3, 5, and 7 are selected for values of $k$, that is, the number of neighbours considered. Discovery of an optimal value for $k$, however, may take considerable time. A classification accuracy of 100% for all data sets is achieved for some but not all of the reducts generated. The process of generating such large numbers of reducts, however, is computationally expensive. The FS approach is compared with two other rough set approaches (Shen & Chouchoulas, 2000) and (Zhong *et al.*, 2001), which also perform well; however, the authors argue that their method performs better on the basis of classification results.

Again, as with the application example in Section 5.2.1, what becomes apparent is the number of tunable parameters. Despite very high classification accuracies being achieved, these are subjective and can influence the final result. The authors mention a parameter for the discretization of the data, another for the FS approach, and of course $k$ for the classification step. Note that if a hybrid fuzzy-rough approach rather than the current rough set approach were to be employed, the discretization step could be eliminated completely. This would also ensure that any potential loss of information would not occur due to the discretization step.

## 6 Conclusion

This paper has presented an overview of RST and its extensions along with representative, theoretical, and practical application examples. In particular, this review has introduced the basic

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

concepts of rough sets: upper and lower approximations, positive, negative, and boundary regions, rough set dependency, and reducts. In order to further develop its potential field of applicability, and to address its theoretical drawbacks in terms of application to real-valued, noisy, and ordinal data, the extensions of RST are also explored. Specifically, tolerance rough sets, fuzzy-rough sets, VPRS, dominance-based rough sets, and vaguely quantified rough sets are described in detail. Other extensions and hybridizations which also extend the traditional RST are also covered.

There are a number of areas of RST, particularly with respect to the hybridization of rough sets which remain to be explored. Fuzzy-rough classification is also an area which has much potential as reflected in Section 5.1.1, as is the application of fuzzy-rough sets to the area of clustering.

It is interesting to note that in terms of rough set classification, and in spite of the level of publication in this area, there has not been a comprehensive and far-reaching review of rough set classification techniques. The reason for this may partly lie in the fact that RST is usually allied to other soft computing methods when used for classification. This paper makes an initial contribution towards such a review of the present work in this area.

The hybridization of rough set extensions also holds some potential. For instance, the marrying of both the DRSM and VQRS would result in a noise-tolerant approach that would potentially have the ability to handle ordinal real-valued data with the advantages of VPRS. The hybridization of the TRSM and VPRS would mean that advantage could be taken of the respective flexibility of both approaches, albeit with two tunable parameters. The absence of tunable parameters (and hence the adherence to the original principles of traditional RST) is one of the most attractive properties of fuzzy-rough sets and the reason why it has gained so much recent attention. This will undoubtedly also be the motivation for driving future research in this area, and it is strongly believed that the fuzzy-rough set approach has much to offer both theoretically and in terms of application to new problem domains.

## Acknowledgements

The authors wish to thank the referees for their invaluable advice in revising this paper.

## References

American College of Radiology. 1998. *Illustrated Breast Imaging Reporting and Data System BIRADS*, 3rd edn. American College of Radiology.

Asharaf, A. & Murty, M. N. 2004. An adaptive rough fuzzy single pass algorithm for clustering large data sets. *Pattern Recognition* **36**(12), 3015–3018.

Baldwin, J. F., Lawry, J. & Martin, T. P. 1997. A mass assignment based ID3 algorithm for decision tree induction. *International Journal of Intelligent Systems* **12**(7), 523–552.

Bao, Z., Han, B. & Wu, S. 2006. A novel clustering algorithm based on variable precision rough-fuzzy sets. In *Proceedings of the International Conference on Intelligent Computing (ICIC 2006)*. Kunming, China, August 16–19, 284–289.

Bazan, J., Nguyen, H. S., Nguyen, S. H., Synak, P. & Wroblewski, J. 2000. Rough set algorithms in classification problem. In *Rough Set Methods and Applications*, Polkowski, L., Tsumoto, S. & Lin, T. Y. (eds). Physica-Verlag, 49–88.

Bell, D. A. & Guan, J. W. 1998. Computational methods for rough classification and discovery. *Journal of the American Society for Information Science* **5**, 403–414.

Beynon, M. J. 2000. An investigation of $\beta$-reduct selection within the variable precision rough sets model. In *Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing (RSCTC 2000)*, Banff, Canada, 114–122.

Beynon, M. J. 2001. Reducts within the variable precision rough sets model: a further investigation. *European Journal of Operational Research* **134**(3), 592–605.

Bezdek, J. C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press.

Bhatt, R. B. & Gopal, M. 2004. FRID: fuzzy-rough interactive dichotomizers. In *Proceeding of the 2004 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'04)*, Budapest, 1337–1342.

Bian, H. & Mazlack, L. 2003. Fuzzy-rough nearest-neighbor classification approach. In *Proceeding of the 22nd International Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, Chicago, USA, 500–505.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

Boixader, D., Jacas, J. & Recasens, J. 2000. Upper and lower approximations of fuzzy sets. *International Journal of General Systems* **29**(4), 555–568.

Brassard, G. & Bratley, P. 1996. *Fundamentals of Algorithms*. Prentice Hall.

Browne, C., Düntsch, I. & Gediga, G. 1998. IRIS revisited, a comparison of discriminant enhanced rough set data analysis. In *Rough Sets in Knowledge Discovery 2: Applications, Case Studies and Software Systems*, Polkowski, L & Skowron, A. (eds). Physica-Verlag, 347–370.

Chen, D., Zhang, W. X., Yeung, D. & Tsang, E. C. C. 2006. Rough approximations on a complete completely distributive lattice with applications to generalized rough sets. *Information Sciences* **176**(13), 1829–1848.

Chimphlee, S., Salim, N., Ngadiman, M. S. B., Chimphlee, W. & Srinoy, S. 2006. Independent component analysis and rough fuzzy based approach to web usage mining. In *Proceedings of the 24th IASTED International Conference on Artificial Intelligence and Applications*, Deved, V, (ed.). International Association of Science and Technology for Development, 422–427. ACTA Press.

Chimphlee, W., Abdullah, A. H., Sap, M. N. M., Srinoy, S. & Chimphlee, S. 2006a. Anomaly-based intrusion detection using fuzzy rough clustering. *International Conference on Hybrid Information Technology (ICHIT'06)* **1**, 329–334.

Chouchoulas, A. & Shen, Q. 2001. Rough set-aided keyword reduction for text categorisation. *Applied Artificial Intelligence* **15**(9), 843–873.

Cordon, O., Gomide, F., Herrera, F., Hoffmann, F. & Magdalena, L. 2001. Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy Sets and Systems* **141**, 5–31.

Cornelis, C., De Cock, M. & Radzikowska, A. 2007. Vaguely quantified rough sets. In *Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC2007)*, Lecture Notes in Artificial Intelligence **4482**, 87–94.

Cornelis, C. & Jensen, R. 2008. A noise-tolerant approach to fuzzy-rough feature selection. In *Proceedings of the 17th International Conference on Fuzzy Systems (FUZZ-IEEE'08)*, Hong Kong, China, 1598–1605.

Dash, M. & Liu, H. 1997. Feature selection for classification. *Intelligent Data Analysis* **1**(3), 131–156.

Davis, M., Logemann, G. & Loveland, D. 1962. A machine program for theorem proving. *Communications of the ACM* **5**, 394–397.

De Cock, M., Cornelis, C. & Kerre, E. E. 2004. Fuzzy rough sets: beyond the obvious. *IEEE International Conference on Fuzzy Systems* **1**, 103–108.

Deogun, J. S., Raghavan, V. V. & Sever, H. 1994. Rough set based classification methods and extended decision tables. In *Proceedings of the International Workshop on Rough Sets and Soft Computing*. San Jose, California, 302–309.

Deogun, J. S., Raghavan, V. V. & Sever, H. 1995. Exploiting upper approximations in the rough set methodology. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*. Quebec, Canada, 1–10.

Devijver, P. & Kittler, J. 1982. *Pattern Recognition: A Statistical Approach*. Prentice Hall.

Dubois, D. & Prade, H. 1990. Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems* **17**, 191–209.

Dubois, D. & Prade, H. 1992. Putting rough sets and fuzzy sets together. In *Intelligent Decision Support: Handbook of Applications and Advances of the Sets Theory*, Slowinski, R. (ed.). Kluwer, 203–232.

Dunn, J. C. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* **3**, 32–57.

Glymin, M. & Ziarko, W. 2007. Rough set approach to spam filter learning. In *Proceedings of Rough Sets and Emerging Intelligent Systems Paradigms (RSEISP'07)*, Lecture Notes in Artificial Intelligence **4585**, 350–359.

Greco, S., Matarazzo, B. & Slowiński, R. 2001. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research* **129**(1), 1–47.

Grzymala-Busse, D. M. & Grzymala-Busse, J. W. 1995. The usefulness of machine learning approach to knowledge acquisition. *Computational Intelligence* **11**, 268–279.

Grzymala-Busse, J. W. & Wang, C. P. B. 1996. Classification methods in rule induction. In *Proceedings of the 5th Intelligent Information Systems Workshop*, Dęblin, Poland, 120–126.

Grzymala-Busse, J. W. 2003. A comparison of three strategies to rule induction from data with numerical attributes. In *Proceedings of the International Workshop on Rough Sets in Knowledge Discovery (RSKD 2003)*, Warsaw, Poland, 132–140.

Grzymala-Busse, J. W. 2006. Rough set theory with applications to data mining. In *Real World Applications of Computational Intelligence, Studies in Fuzziness and Soft Computing Series*, Negoita, M. & Reusch, B. (eds). Springer, Heidelberg, 223–244.

Han, J., Hu, X. & Lin, T. Y. 2005. Feature subset selection based on relative dependency between Attributes. In *Rough Sets and Current Trends in Computing: 4th International Conference (RSCTC 2004)*. Uppsala, Sweden, June 1–5, 176–185.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

Heath, M., Bowyer, K., Kopans, D., Moore, R. & Kegelmeyer, P. J. 2000. The digital database for screening mammography. In *Proceedings of the International Workshop on Digital Mammography*, Madison, Wiscowsow, USA, 212–218.

Hirano, S. & Tsumoto, S. 2000. Rough clustering and its application to medicine. *Journal of Information Sciences* **124**, 125–137.

Hirano, S. & Tsumoto, S. 2003. Indiscernibility-based clustering: rough clustering. In *International Fuzzy Systems Association World Congress*, Lecture Notes in Computer Science **2715**, 378–386. Springer-Verlag.

Ho, B. & Nguyen, N. B. 2002. Nonhierarchical document clustering based on a tolerance rough set model. *International Journal of Intelligent Systems* **17**(2), 199–212.

Ho, T. B., Kawasaki, S. & Nguyen, N. B. 2006. Documents clustering using tolerance rough set model and its application to information retrieval. In *Studies In Fuzziness and Soft Computing, Intelligent Exploration of the Web*, Szczepaniak, P.S., Segovia, J., Karprzyk, J., & Zadeh, L.A. (eds). Physica-Verlag, Heidelberg, 181–196.

Hong, T. P., Liou, Y. L. & Wang, S. L. 2006. Learning with hierarchical quantitative attributes by fuzzy rough sets. In *Proceedings of the 2006 Joint Conference on Information Sciences, Advances in Intelligent Systems Research*, Taiwan, ROC, 1309–1312.

Hu, Q.-H. & Yu, D.-R. 2004. Variable precision dominance based rough set model and reduction algorithm for preference-ordered data. *Proceedings of the 2004 International Conference on Machine Learning and Cybernetics* **4**, 2279–2284.

Hu, Q.-H. & Yu, D.-R. 2005. Fuzzy rough C-means clustering. In *World Congress on Fuzzy Logic, Soft Computing, Computational Intelligence: Theories and Applications (IFSA2005)*, Springer Lecture Notes, Tsinghua, Beijing.

Hu, Q., Yu, D. & Xie, Z. 2006. Information-preserving hybrid data reduction based on fuzzy-rough techniques. *Pattern Recognition Letters* **27**(5), 414–423.

Hu, Q., Zhao, H., Xie, Z. & Yu, D. 2007a. Consistency based attribute reduction. In *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining 2007*, Zhou, Z., Li, H. & Yang, Q. (eds). Lecture Notes in Computer Science **4426**, 96–107.

Hu, Q., Xie, Z. & Yu, D. 2007b. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. *Pattern Recognition* **40**, 3509–3521.

Inuiguchi, M. & Tsurumi, M. 2006. Measures based on upper approximations of rough sets for analysis of attribute importance and interaction. *International Journal of Innovative Computing Information and Control* **2**(1), 1–12.

Jelonek, J., Krawiec, K., Slowiński, R., Stefanowski, J. & Szymas, J. 1994. Rough sets as an intelligent front-end for the neural network. In *Proceedings of the First National Conference on Neural Networks Their Applications 2*, Poland, 116–122.

Jensen, R. & Shen, Q. 2004a. Semantics-preserving dimensionality reduction: rough and fuzzy-rough based approaches. *IEEE Transactions on Knowledge and Data Engineering* **16**(12), 1457–1471.

Jensen, R. & Shen, Q. 2004b. Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets and Systems* **141**(3), 469–485.

Jensen, R. & Shen, Q. 2005. Fuzzy-rough data reduction with ant colony optimization. *Fuzzy Sets and Systems* **149**(1), 5–20.

Jensen, R. & Shen, Q. 2007. Fuzzy-rough sets assisted attribute selection. *IEEE Transactions on Fuzzy Systems* **15**(1), 73–89.

Jensen, R. & Cornelis, C. 2008. A new approach to fuzzy-rough nearest neighbour classification. In *Proceedings of the 6th International Conference on Rough Sets and Current Trends in Computing*, Akron, Ohio, USA, 310–319.

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

Jensen, R. & Shen, Q. 2009. New approaches to fuzzy-rough feature selection. *IEEE Transactions on Fuzzy Systems* **17**(4), 824–838.

Jian, L.-R. & Li, M.-Y. 2007. An extension of VPRS model based on dominance relation. *Proceedings of the 4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)* **3**, 113–118.

Kawasaki, S., Nguyen, N. B. & Ho, T. B. 2000. Hierarchical document clustering based on tolerance rough set model. In *Principles of Data Mining and Knowledge Discovery, 4th European Conference (PKDD 2000)*, Lyon, France (September 13–16, 2000), Zighed, D. A., Komorowski, H. J. & Zytkow, J. M. (eds). Lecture Notes in Computer Science **1910**, 13–27. Springer.

Ke, L., Feng, Z. & Ren, Z. 2008. An efficient ant colony optimization approach to attribute reduction in rough set theory. *Pattern Recognition Letters* **29**, 1351–1357.

Keller, J. M., Gray, M. R. & Givens, J. A. 1985. A fuzzy K-nearest neighbor algorithm. *IEEE Transactions on Systems Man and Cybernetics* **15**(4), 580–585.

Kim, D. & Bang, S.-Y. 2000. A handwritten numeral character classification using tolerant rough set. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(9), 923–937.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

Komorowski, J., Pawlak, Z., Polkowski, L. & Skowron, A. 1999. Rough sets: a tutorial. In *Rough-Fuzzy Hybridization: A New Trend in Decision Making*, Pal, S. K. & Skowron, A. (eds). Springer-Verlag, 3–98.

Kotłowski, W., Dembczyński, K., Greco, S. & Słowiński, R. 2008. Stochastic dominance-based rough set model for ordinal classification. *International Journal of Information Sciences* **178**(21), 4019–4037.

Kryszkiewicz, M. 1994. *Maintenance of Reducts in the Variable Precision Rough Sets Model*. ICS Research Report 31/94, Warsaw University of Technology.

Kumar, P., Krishna, P. R., Bapi, R. S. & De, S. K. 2007. Rough clustering of sequential data. *Data & Knowledge Engineering* **63**(2), 183–199.

Li, R. & Wang, Z.-O. 2004. Mining classification rules using rough sets and neural networks. *European Journal of Operational Research* **157**, 439–448.

Li, M., Wu, C., Zhang, Y. & Yue, Y. 2006. An improved BP network classifier based on VPRS feature reduction. *The Sixth World Congress on Intelligent Control and Automation (WCICA 2006)* **2**, 9677–9680.

Lingras, P. 1996. Rough neural networks. *Proceedings of the Sixth International Conference on Information Processing Management of Uncertainty in Knowledge-based Systems (IPMU'96)* **2**, 1445–1450.

Lingras, P. 1997. Comparison of neofuzzy rough neural networks. In *Proceedings of the Fifth International Workshop on Rough Sets Soft Computing (RSSC'97)*, 259–262.

Lingras, P. & Davies, C. 2001. Applications of rough genetic algorithms. *Computational Intelligence* **17**(3), 435–445.

Lingras, P. & West, C. 2004. Interval set clustering of web users with rough K-means. *Journal of Intelligent Information Systems* **23**(1), 5–16.

Lingras, P., Hogo, M. & Snorek, M. 2004. Interval set clustering of web users using modified Kohonen self-organizing maps based on the properties of rough sets. *Web Intelligence and Agent Systems* **2**(3), 217–230.

Mak, B. & Munakata, T. 2002. Rule extraction from expert heuristics: a comparative study of rough sets with neural networks and ID3. *European Journal of Operational Research* **136**, 212–229.

Malyszko, D. & Stepaniuk, J. 2008. Standard and fuzzy rough entropy clustering algorithms in image segmentation. *Rough Sets and Current Trends in Computing* **5306**, 409–418.

Mac Parthaláin, N., Shen, Q. & Jensen, R. 2010. A distance measure approach to exploring the rough set boundary region for attribute reduction. *IEEE Transactions on Knowledge and Data Engineering* **22**(3), 306–317.

Mac Parthaláin, N., Jensen, R., Shen, Q. & Zwiggelaar, R. 2010. Rough and fuzzy-rough methods for mammographic data analysis. *Intelligent Data Analysis—An International Journal* **14**(2), 225–244.

Mac Parthaláin, N. & Shen, Q. 2009. Exploring the boundary region of tolerance rough sets for feature selection. *Pattern Recognition* **42**(5), 655–667.

McKee, T. & Lensberg, T. 2002. Genetic programming and rough sets: a hybrid approach to bankruptcy classification. *European Journal of Operational Research* **140**(2), 436–451.

Mi, J. S. & Zhang, W. X. 2004. An axiomatic characterization of a fuzzy generalization of rough sets. *Information Sciences* **160**(1–4), 235–249.

Mieszkowicz-Rolka, A. & Rolka, L. 2004. Fuzzy implication operators in variable precision fuzzy rough sets model. *Lecture Notes in Computer Science (LNCS)* **3070**, Springer, Heidelberg, 498–503.

Mitra, S. & Banerjee, M. 1996. Knowledge based neural net with rough sets. In *Methodologies for the Conception, Design, Application of Intelligent Systems, Proceedings of the Fourth International Conference on Soft Computing (IIZUKA'96)*, Yamakawa, T. & Matsumoto, G. (eds). World Scientific, 213–216.

Mitra, P. & Mitra, S. 2000. Staging of cervical cancer with soft computing. *IEEE Transactions on Biomedical Engineering* **47**(7), 934–940.

Modrzejewski, M. 1993. Feature selection using rough sets theory. In *Proceedings of the 11th International Conference on Machine Learning*, New Brunswick, NJ, USA, 213–226.

Molina, L. C., Belanche, L. & Nebot, A. 2002. Feature selection algorithms: a survey and experimental evaluation. In *Proceedings of ICDM02*, Maebashi City, Japan, 306–313.

Momin, B. F., Mitra, S. & Gupta, R. D. 2006. Reduct generation and classification of gene expression data. *Proceedings of the 2006 International Conference on Hybrid information Technology (ICHIT06)* **1**, 699–708.

Morsi, N. N. & Yakout, M. M. 1998. Axiomatics for fuzzy rough sets. *Fuzzy Sets and Systems* **100**(1–3), 327–342.

Narendra, P. & Fukunaga, K. 1977. A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers* **C-26**(9), 917–922.

Ngo, C. L. & Nguyen, H. S. 2004. A tolerance rough set approach to clustering web search results. In *Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases* (Pisa, Italy, September 20–24, 2004), Boulicaut, J., Esposito, F., Giannotti, F. & Pedreschi, D. (eds). Lecture Notes in Computer Science **3202**, Springer-Verlag New York, New York, 515–517.

Nguyen, S. H. & Skowron, A. 1997a. Searching for relational patterns in data. In *Proceedings of the First European Symposium on Principles of Data Mining and Knowledge Discovery*, Trondheim, Norway, 265–276.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

Nguyen, S. H. & Slezak, D. 2004. Approximate reducts and association rules correspondence and complexity results. *Lecture Notes in Computer Science (LNCS)*, Zhong, N., Skowron, A., & Ohsuga, S. (eds). **1711**, Springer, Heidelberg, 137–145.

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

Oliver, A., Freixenet, J., Marti, R., Pont, J., Perez, E., Denton, E. R. E. & Zwiggelaar, R. 2008. A novel breast tissue density classification methodology. *IEEE Transactions on Information Technology in Biomedicine* **12**(1), 55–65.

Pal, S. K. 2004. *Pattern Recognition Algorithms for Data Mining*. Chapman and Hall.

Pattaraintakorn, P. & Cercone, N. 2007. Integrating rough set theory and medical applications. *Applied Mathematics Letters* **21**(4), 400–403.

Pawlak, Z. 1982. Rough sets. *International Journal of Computing and Information Sciences* **11**, 341–356.

Pawlak, Z. 1984. Rough classification. *International Journal of Man-Machine Studies* **20**, 469–483.

Pawlak, Z., Slowinski, K. & Slowinski, R. 1986. Rough classification of patients after highly selective vagotomy for duodenal ulcer. *International Journal of Man Machine Studies* **24**, 413–433.

Pawlak, Z. 1991. *Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishing.

Pawlak, Z. & Skowron, A. 1993. A rough set approach for decision rules generation. ICS Research Report 23/93, Warsaw University of Technology. *Proceedings of the International Joint Conference on Artificial Intelligence '93 Workshop W12: The Management of Uncertainty in AI*, France.

Pawlak, Z. 2003. Some issues on rough sets. *Lecture Notes on Computer Science*, *Transactions on Rough Sets* **1**, Springer, Heidelberg, 1–53.

Pedrycz, W. 1999. Shadowed sets: bridging fuzzy and rough sets. In *Rough-Fuzzy Hyridisation*, Pal, S. K. & Skowron, A. (eds). Springer-Verlag, 179–199.

Peters, J. F., Skowron, A., Suraj, Z., Rzasa, W. & Bokowski, M. 2002. Clustering: a rough set approach to constructing information granules. In *Proceedings of the 6th International Conference on Soft Computing and Distributed Processing*, Rzeszow, Poland, 57–61.

Petrosino, A. & Ceccarelli, M. 2000. Unsupervised texture discrimination based on rough fuzzy sets and parallel hierarchical clustering. In *Proceedings of the International Conference on Pattern Recognition (ICPR '00)* **3** (September 03–08, 2000), IEEE Computer Society, Washington, DC.

Polkowski, L. & Skowron, A. 1998. Rough sets: a perspective. In *Rough Sets in Knowledge Discovery 1: Methodology and Applications*, Polkowski, L. & Skowron, A. (eds). Physica-Verlag, 31–56.

Qina, K. & Pei, Z. 2005. On the topological properties of fuzzy rough sets. *Fuzzy Sets and Systems* **151**(3), 601–613.

Radzikowska, A. M. & Kerre, E. E. 2002. A comparative study of fuzzy rough sets. *Fuzzy Sets and Systems* **126**(2), 137–155. Elsevier, Amsterdam.

Radzikowska, A. M. & Kerre, E. E. 2004. Fuzzy rough sets based on residuated lattices. *Transactions on Rough Sets II, Lecture Notes in Computer Science (LNCS)* **3135**, 278–296.

Sarkar, M. 2000. Fuzzy-rough nearest neighbors algorithm. In *Proceedings of the IEEE Conference on Systems Man and Cybernetics*, Nashville, TN, USA, 3556–3561.

Sarkar, M. 2007. Fuzzy-rough nearest neighbors algorithm. *Fuzzy Sets and Systems* **158**, 2123–2152.

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

Shan, D., Ishii, N., Hujun, Y., Allinson, N., Freeman, R., Keane, J. & Hubbard, S. 2002. Feature weights determining of pattern classification by using a rough genetic algorithm with fuzzy similarity measure. In *Proceedings of Intelligent Data Engineering and Automated Learning*, Manchester, UK, 544–550.

Shang, C. & Shen, Q. 2002. Rough feature selection for neural network based image classification. *International Journal of Image and Graphics* **2**(4), 541–555.

Shao, M.-W. & Zhang, W.-X. 2004. Dominance relation and rules in an incomplete ordered information system. *International Journal of Intelligent Systems* **20**(1), 13–20.

Shen, Q. & Chouchoulas, A. 2000. 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.

Shen, Q. & Chouchoulas, A. 2002. A rough-fuzzy approach for generating classification rules. *Pattern Recognition* **35**(2), 2425–2438.

Shen, Q. & Jensen, R. 2004. Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. *Pattern Recognition* **37**(7), 1351–1363.

Skowron, A. & Rauszer, C. 1992. The discernibility matrices and functions in information systems. In *Intelligent Decision Support: Handbook of Applications and Advances to Rough Sets Theory*, Slowinski, R. (ed.). Kluwer Academic, 331–362.

Skowron, A. 1993. Boolean reasoning for decision rules generation. In *Proceedings of the 7th International Symposium ISMIS'93*, Komorowski, J. & Ras, Z. (eds). Lecture Notes in Artificial Intelligence, Trondheim, Norway **689**, 295–305. Springer-Verlag.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

Skowron, A. & Stepaniuk, J. 1994. Generalized approximation spaces. In *Proceedings of the 3rd International Workshop on Rough Sets and Soft Computing*, San Jose, California, USA, 156–163.

Skowron, A. & Stepaniuk, J. 1996. Tolerance approximation spaces. *Fundamenta Informaticae* **27**, 245–253.

Skowron, A., Pawlak, Z., Komorowski, J. & Polkowski, L. 2002. A rough set perspective on data and knowledge. In *Handbook of Data Mining and Knowledge Discovery*, Kloesgen, W., & Zytkow, J. (eds). Oxford University Press, 134–149.

Slowinski, R. & Vanderpooten, D. 1997. Similarity relations as a basis for rough approximations. In *Advances in Machine Intelligence and Soft Computing*, Wang, P. P. (ed.). Bookwrights, 17–33.

Slowinski, R. & Vanderpooten, D. 2000. A generalized definition of rough approximations based on similarity. *IEEE Transactions on Knowledge and Data Engineering* **12**(2), 331–336.

Slowinski, K., Stefanowski, J. & Siwinski, D. 2002. Application of rule induction and rough sets to verification of magnetic resonance diagnosis. *Fundamenta Informaticae* **53**(3/4), 345–363.

Srinivasan, P., Ruiz, M. E., Kraft, D. H. & Chen, J. 1998. Vocabulary mining for information retrieval: rough sets and fuzzy sets. *Information Processing & Management* **37**(1), 15–38.

Stefanowski, J. 1998. On rough set based approaches to induction of decision rules. In *Rough Sets in Knowledge Discovery*, Skowron, A. & Polkowski, L. (eds). **1**, Physica-Verlag, 500–529.

Suckling, J., Partner, J., Dance, D. R., Astley, S. M., Hutt, I., Boggis, C. R. M., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S. L., Taylor, Betal, P. & Savage, J. 1994. The mammographic image analysis society digital mammogram database. In *International Workshop on Digital Mammography*, York, UK, 211–221.

Swiniarski, R., Hunt, F., Chalvet, D. & Pearson, D. 1995. Intelligent data processing and dynamic process discovery using rough sets, statistical reasoning and neural networks in a highly automated production system. In *Proceedings of the First European Conference on Application of Neural Networks in Industry*, Helsinki, Finland.

Swiniarski, R. 1998. Rough sets Bayesian methods applied to cancer detection. In *Proceeding of the First International Conference on Rough Sets and Soft Computing (RSCTC'98)*, Polkowski, L. & Skowron, A. (eds). LNAI **1424**, 609–616. Springer-Verlag, 275–300.

Swiniarski, R. 1999. Rough sets and principal component analysis and their applications in data model building and classification. In *Rough Fuzzy Hybridization: New Trends in Decision Making*, Pal, S. K. & Skowron, A. (eds). Springer-Verlag, 275–300.

Swiniarski, R. & Skowron, A. 2003. Rough set methods in feature selection and recognition. *Pattern Recognition Letters* **24**(6), 83–849.

Thangavel, K., Pethalakshmi, A. & Jaganathan, P. 2006. A comparative analysis of feature selection algorithms based on rough set theory. *International Journal of Soft Computing* **1**(4), 288–294.

Thiele, H. 1998. *Fuzzy Rough Sets Versus Rough Fuzzy Sets – An Interpretation and a Comparative Study Using Concepts of Modal Logics*. Technical report no. CI-30/98, University of Dortmund.

Tsang, E. C. C., Chen, D., Yeung, D. S., Wang, X.-Z. & Lee, J. 2008. Attributes reduction using fuzzy rough sets. *IEEE Transactions on Fuzzy Systems* **16**(5), 1130–1141.

Wang, J. & Wang, J. 2001. Reduction algorithms based on discernibility matrix: the ordered attributes method. *Journal of Computer Science & Technology* **16**(6), 489–504.

Wang, Y., Ding, M., Zhou, C. & Zhang, T. 2005. A hybrid method for relevance feedback in image retrieval using rough sets and neural networks. *International Journal of Computational Cognition* **3**(1), 78–87.

Wang, X., Yang, J., Teng, X. & Peng, N. 2005a. Fuzzy-rough set based nearest neighbor clustering classification algorithm. *Lecture Notes in Computer Science* **3613**, 370–373.

Wang, Z., Shao, X., Zhang, G. & Zhu, H. 2005b. Integration of variable precision rough set and fuzzy clustering: an application to knowledge acquisition for manufacturing process planning. In *Proceedings of the 10th Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing (RSFDGrC 2005)*, Regina, Canada.

Wang, X., Yang, J., Teng, X., Xia, W. & Jensen, R. 2007. Feature selection based on rough sets and particle swarm optimization. *Pattern Recognition Letters* **28**(4), 459–471.

Wojdyllo, P. 1998. Wavelets, rough sets arificial neural networks in EEG analysis. In *Proceedings of the First International Conference on Rough Sets and Soft Computing (RSCTC'98)*, Polkowski, L. & Skowron, A. (eds). LNAI **1424**, 444–449. Springer-Verlag.

Wróblewski, J. 1995. Finding minimal reducts using genetic algorithms. In *Proceedings of the International Workshop on Rough Sets Soft Computing at Second Annual Joint Conference on Information Sciences (JCIS'95)*, Wrightsville Beach, NC, USA, 186–189.

Wu, W. Z., Mi, J. S. & Zhang, W. X. 2003. Generalized fuzzy rough sets. *Information Sciences* **151**, 263–282.

Wu, W. Z. & Zhang, W. X. 2004. Constructive and axiomatic approaches of fuzzy approximation operators. *Information Sciences* **159**(3–4), 233–254.

Wu, W. Z. 2005. A study on relationship between fuzzy rough approximation operators and fuzzy topological spaces. In *Fuzzy Systems and Knowledge Discovery 2005*, Wang, L. & Jin, Y. (eds). Lecture Notes in Artficial Intelligence **3613**, 167–174. Springer, Heidelberg.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press

Wu, W. Z., Leung, Y. & Mi, J. S. 2005. On characterizations of (I,T)-fuzzy rough approximation operators. *Fuzzy Sets and Systems* **154**(1), 76–102.

Wygralak, M. 1989. Rough sets and fuzzy sets – some remarks on interrelations. *Fuzzy Sets and Systems* **29**(2), 241–243.

Yahia, M., Mahmod, R., Sulaiman, N. & Ahmad, F. 2000. Rough neural expert systems. *Expert Systems with Applications* **27**(2), 87–99.

Yao, Y. Y. 1997. Combination of rough and fuzzy sets based on $\alpha$-level sets. In *Rough Sets and Data Mining: Analysis of Imprecise Data*, Lin, T. Y. & Cereone, N. (eds). Kluwer Academic Publishers, 301–321.

Yeung, D. S., Chen, D., Tsang, E. C. C., Lee, J. W. T. & Xizhao, W. 2005. On the generalization of fuzzy rough sets. *IEEE Transactions on Fuzzy Systems* **13**(3), 343–361.

Yi, G., Hu, H. & Lu, Z. 2005. Web document classification based on extended rough set. In *Proceedings of the Sixth International Conference on Parallel and Distributed Computing Applications and Technology (PDCAT)*, IEEE Computer Society, Washington, DC, 916–919.

Yun, O. & Ma, J. 2006. Land cover classification based on tolerant rough set. *International Journal of Remote Sensing* **27**(14), 3041–3047.

Zadeh, L. A. 1965. Fuzzy sets. *Information and Control* **8**(3), 338–353.

Zhao, Y., Zhang, H. & Pan, Q. 2003. Classification using the variable precision rough set. In *Proceedings of Rough Sets, Fuzzy Sets, Data Mining and Granular Computing 2003* **2639**, 350–353. Chongqing.

Zhao, S. & Zhang, Z. 2005. A generalized definition of rough approximation based on similarity in variable precision rough sets. In *Proceedings of the 2005 International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 3153–3156.

Zhao, Y., Zhou, X. & Tang, G. 2005. A rough set-based fuzzy clustering. In *Proceedings of the Second Asia Information Retrieval Symposium*, Jeju Island, Korea, 401–409.

Zhao, W. Q. & Zhu, Y. L. 2006. Classifying email using variable precision rough set approach. *Lecture Notes in Artificial Intelligence* **4062**, 766–771.

Zheng, X. & Wang, J. 2008. Power transformer fault diagnosis based on variable precision rough set. In *Proceedings of the 3rd International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, Nanjing, China, 1353–1358.

Zhong, N., Dong, J. & Ohsuga, S. 2001. Using rough sets with heuristics for feature selection. *Journal of Intelligent Information Systems* **16**(3), 199–214.

Ziarko, W. 1993. Variable precision rough set model. *Journal of Computer and Systems Sciences* **46**(1), 39–59.

Ziarko, W. 2003. Acquisition of hierarchy-structured probabilistic decision tables and rules from data. *Expert Systems* **20**(5), 305–310.

https://doi.org/10.1017/S0269888910000263 Published online by Cambridge University Press