International Journal of Automation and Computing 04(3), July 2007, 217-228
DOI: 10.1007/s11633-007-0217-y

Rough Sets, Their Extensions and Applications

Qiang Shen* Richard Jensen

Department of Computer Science, The University of Wales, Aberystwyth, Ceredigion SY23 3DB, UK

Abstract: Rough set theory provides a useful mathematical foundation for developing automated computational systems that can help understand and make use of imperfect knowledge. Despite its recency, the theory and its extensions have been widely applied to many problems, including decision analysis, data mining, intelligent control and pattern recognition. This paper presents an outline of the basic concepts of rough sets and their major extensions, covering variable precision, tolerance and fuzzy rough sets. It also shows the diversity of successful applications these theories have entailed, ranging from financial and business, through biological and medicine, to physical, art, and meteorological.

Keywords: Rough sets, data processing, fuzzy sets.

1 Introduction

Dealing with incomplete or imperfect knowledge is the core of much research in computational intelligence and cognitive sciences. Being able to understand and manipulate such knowledge is of fundamental significance to many theoretical developments and practical applications of automation and computing, especially in the areas of decision analysis, machine learning and data mining, intelligent control and pattern recognition. Rough set theory[1, 2] offers one of the most distinct and recent approaches for this.

Indeed, since its invention, this theory has been successfully utilised to devise mathematically sound and often, computationally efficient techniques for addressing problems such as hidden pattern discovery from data, data reduction, data significance evaluation, decision rule generation, and data-driven inference interpretation[3]. Owing to the recognition of the existing and potential important impact of this theory, it has attracted world-wide attention of further research and development, resulting in various extensions to the original theory and increasingly widening fields of application. This paper attempts to offer a concise overview of the basic ideas of rough set theory, and its major extensions with sample applications. Further details can be found in the literature (e.g. *LNCS Transactions on Rough Sets*) and on the internet (e.g. www.roughsets.org).

This paper is organized as follows. The next section outlines the preliminary concepts of rough set theory. Section 3 introduces three major extensions made to the original theory, covering variable precision rough sets, tolerance rough sets and fuzzy rough sets. Section 4 describes a range of practical applications of the theory and a specific theoretical development on feature selection that is based on the use of the original or extended rough set theories and that is itself an important topic in automation and computing. The final section concludes the paper and points out some interesting further work.

Manuscript received date March 5, 2007; revised date May 14, 2007
This work was partly supported by the UK EPSRC Grant(No. GR/S98603/01).
*Corresponding author. E-mail address: qqs@aber.ac.uk

2 Rough set theory

Rough set theory (RST)is an extension of conventional set theory that supports approximations in decision making[1,4−7]. It possesses many features in common (to a certain extent) with the Dempster-Shafer theory of evidence[8] and fuzzy set theory[9, 10]. A rough set is itself 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 subset of interest, whereas the upper approximation is a description of the objects which possibly belong to the subset.

It works by exploring and exploiting the granularity structure of the data only. This is a major difference when compared with Dempster-Shafer theory[11, 12] and fuzzy set theory[13] which require probability assignments and membership values respectively. However, this does not mean that *no* model assumptions are made. In fact by using only the given information, the theory assumes that the data is a true and accurate reflection of the real world (which may not be the case). The numerical and other contextual aspects of the data are ignored which may seem to be a significant omission, but keeps model assumptions to a minimum.

2.1 Information and decision systems

An information system can be viewed as a table of data, consisting of objects (rows in the table) and attributes (columns). In medical datasets, for example, patients might be represented as objects and measurements such as blood pressure, form attributes. The attribute values for a particular patient is their specific reading for that measurement. Throughout this paper, the terms attribute, feature and variable are used interchangeably.

An information system may be extended by the inclusion of decision attributes. Such a system is termed a decision system. For example, the medical information system mentioned previously could be extended to include patient classification information, such as whether a patient is ill or healthy. A more abstract example of a decision

system can be found in Table 1. Here, the table consists of four conditional features (a, b, c, d), a decision feature (e) and eight objects. A decision system is *consistent* if for every set of objects whose attribute values are the same, the corresponding decision attributes are identical.

Table 1 An example dataset

$x \in \mathbb{U}$	a	b	c	d	$\Rightarrow$	e
0	S	R	T	T		R
1	R	S	S	S		T
2	T	R	R	S		S
3	S	S	R	T		T
4	S	R	T	R		S
5	T	T	R	S		S
6	T	S	S	S		T
7	R	S	S	R		S

More formally, $I = (\mathbb{U}, \mathbb{A})$ is 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 decision systems, $\mathbb{A} = \{\mathbb{C} \cup \mathbb{D}\}$ where $\mathbb{C}$ is the set of input features and $\mathbb{D}$ is the set of class indices. Here, a class index $d \in \mathbb{D}$ is itself a variable $d : \mathbb{U} \rightarrow \{0, 1\}$ such that for $a \in \mathbb{U}, d(a) = 1$ if a has class d and $d(a) = 0$ otherwise.

2.2 Indiscernibility

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

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

Note that this corresponds to the equivalence relation for which two objects are equivalent if and only if they have the same vectors of attribute values for the attributes in P. The partition of $\mathbb{U}$, determined by *IND(P)* is denoted $\mathbb{U}/IND(P)$ or $\mathbb{U}/P$, which is simply the set of equivalence classes generated by $IND(P)$:

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

where

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

If $(x, y) \in IND(P)$, then x and y are indiscernible by attributes from P. The equivalence classes of the indiscernibility relation with respect to P are denoted $[x]_P$, $x \in \mathbb{U}$. For the illustrative example, if $P = \{b,c\}$, then objects 1, 6 and 7 are indiscernible; as are objects 0 and 4. *IND(P)* creates the following partition of $\mathbb{U}$:

$$\mathbb{U}/IND(P) = \mathbb{U}/IND(b) \otimes \mathbb{U}/IND(c) = \{\{0, 2, 4\}, \{1, 3, 6, 7\}, \{5\}\} \otimes \{\{2, 3, 5\}, \{1, 6, 7\}, \{0, 4\}\} = \{\{2\}, \{0, 4\}, \{3\}, \{1, 6, 7\}, \{5\}\}.$$

2.3 Lower and upper approximations

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 the classical crisp set X:

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

$$\overline{P}X = \{x \mid [x]_P \cap X \neq \emptyset\}. \quad (5)$$

It is such a tuple $\langle \underline{P}X, \overline{P}X \rangle$ that is called a rough set. Consider the approximation of concept X in Fig. 1 Each square in the diagram represents an equivalence class, generated by indiscernibility between object values. Using the features in set B, *via* these equivalence classes, the lower and upper approximations of X can be constructed. Equivalence classes contained within X belong to the lower approximation. Objects lying within this region can be said to belong definitely to concept X. Equivalence classes within X and along its border form the upper approximation. Those objects in this region can only be said to possibly belong to the concept.

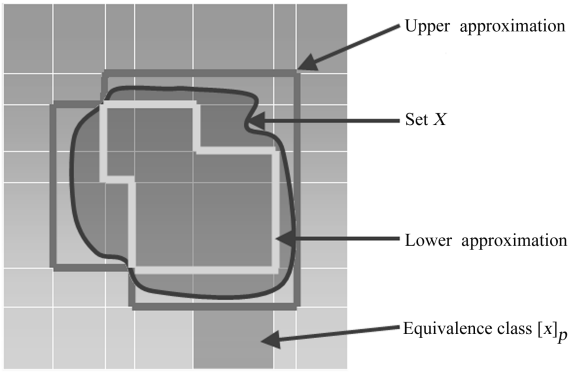

Fig. 1 A rough set

2.4 Positive, negative and boundary regions

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

$$POS_P(Q) = \bigcup_{X \in \mathbb{U}/Q} \underline{P}X \quad (6)$$

$$NEG_P(Q) = \mathbb{U} - \bigcup_{X \in \mathbb{U}/Q} \overline{P}X \quad (7)$$

$$BND_P(Q) = \bigcup_{X \in \mathbb{U}/Q} \overline{P}X - \bigcup_{X \in \mathbb{U}/Q} \underline{P}X. \quad (8)$$

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

For example, let $P = \{b,c\}$ and $Q = \{e\}$, then

$$POS_P(Q) = \bigcup\{\emptyset, \{2,5\}, \{3\}\} = \{2,3,5\}$$
$$NEG_P(Q) = \mathbb{U} - \bigcup\{\{0,4\}, \{2,0,4,1,6,7,5\}, \{3,1,6,7\}\} = \emptyset$$
$$BND_P(Q) = \bigcup\{\{0,4\}, \{2,0,4,1,6,7,5\}, \{3,1,6,7\}\} - \{2,3,5\} = \{0,1,4,6,7\}.$$

This means that objects 2, 3 and 5 can certainly be classified as belonging to a class in attribute e, when considering attributes b and c. The rest of the objects cannot be classified as the information that would make them discernible is absent.

2.5 Attribute dependency and significance

An important issue in data analysis is discovering dependencies between attributes. Intuitively, a set of attributes Q depends totally on a set of attributes P, denoted $P \Rightarrow Q$, if all attribute values from Q are uniquely determined by values of attributes from P. In rough set theory, dependency is defined in the following way:

For P, $Q \subset \mathbb{A}$, it is said that Q depends on P in a degree k $(0 \leq k \leq 1)$, denoted $P \Rightarrow_k Q$, if

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|\mathbb{U}|} \tag{9}$$

where $|S|$ stands for the cardinality of set S. If $k = 1$, Q depends totally on P, if $0 < k < 1$, Q depends partially (in a degree k) on P, and if $k = 0$ then Q does not depend on P. In the example, the degree of dependency of attribute $\{e\}$ from the attributes $\{b,c\}$ is

$$\gamma_{\{b,c\}}(\{e\}) = \frac{|POS_{\{b,c\}}(\{e\})|}{|\mathbb{U}|} = \frac{|\{2,3,5\}|}{|\{0,1,2,3,4,5,6,7\}|} = \frac{3}{8}.$$

By calculating the change in dependency when an attribute is removed from the set of considered possible attributes, an estimate of the significance of that attribute can be obtained. The higher the change in dependency, the more significant the attribute is. If the significance is 0, then the attribute is dispensable without losing information. More formally, given P, Q and an attribute $x \in P$, the significance of attribute x upon Q is defined by

$$\sigma_P(Q,a) = \gamma_P(Q) - \gamma_{P-\{a\}}(Q). \tag{10}$$

For example, if $P = \{a,b,c\}$ and $Q = \{e\}$ then

$$\gamma_{\{a,b,c\}}(\{e\}) = |\{2,3,5,6\}|/8 = 4/8$$
$$\gamma_{\{a,b\}}(\{e\}) = |\{2,3,5,6\}|/8 = 4/8$$
$$\gamma_{\{b,c\}}(\{e\}) = |\{2,3,5\}|/8 = 3/8$$
$$\gamma_{\{a,c\}}(\{e\}) = |\{2,3,5,6\}|/8 = 4/8.$$

And calculating the significance of the three attributes gives

$$\sigma_P(Q,a) = \gamma_{\{a,b,c\}}(\{e\}) - \gamma_{\{b,c\}}(\{e\}) = 1/8$$
$$\sigma_P(Q,b) = \gamma_{\{a,b,c\}}(\{e\}) - \gamma_{\{a,c\}}(\{e\}) = 0$$
$$\sigma_P(Q,c) = \gamma_{\{a,b,c\}}(\{e\}) - \gamma_{\{a,b\}}(\{e\}) = 0.$$

From this it follows that attribute a is indispensable, but attributes b and c can be dispensed with when considering the dependency between the decision attribute and the given individual conditional attributes.

2.6 Reducts

For many application problems, it is often necessary to maintain a concise form of the information system. One way to implement this is to search for a minimal representation of the original dataset. For this, the concept of a *reduct* is introduced and defined as a minimal subset R of the initial attribute set $\mathbb{C}$ such that for a given set of attributes D, $\gamma_R(D) = \gamma_{\mathbb{C}}(D)$. From the literature, R is a minimal subset if $\gamma_{R-\{a\}}(D) \neq \gamma_R(D)$ for all $a \in R$. This means that no attributes can be removed from the subset without affecting the dependency degree. Hence, a minimal subset by this definition may not be the *global* minimum (a reduct of smallest cardinality). A given dataset may have many reduct sets, and the collection of all reducts is denoted by

$$R_{\text{all}} = \{X \mid X \subseteq \mathbb{C}, \gamma_X(D) = \gamma_{\mathbb{C}}(D); \gamma_{X-\{a\}}(D) \neq \gamma_X(D), \forall a \in X\}. \tag{11}$$

The intersection of all the sets in R_{all} is called the *core*, the elements of which are those attributes that cannot be eliminated without introducing more contradictions to the representation of the dataset. For many tasks (for example, feature selection[14]), a reduct of minimal cardinality is ideally searched for. That is, an attempt is to be made to locate a single element of the reduct set $R_{\min} \subseteq R_{\text{all}}$:

$$R_{\min} = \{X \mid X \in R_{\text{all}}, \forall Y \in R_{\text{all}}, |X| \leq |Y|\}. \tag{12}$$

The problem of finding a reduct of an information system has been the subject of much research[15, 16]. The QuickReduct algorithm given in follows (adapted from [17]), attempts to calculate reducts for a decision problem (though the underlying approach can be applied to other tasks), without exhaustively generating all possible subsets:

The QuickReduct algorithm

```
QuickReduct(ℂ,𝔻).
ℂ: the set of all conditional attributes;
𝔻: the set of decision attributes.
1)  R ← {}
2)  do
3)      T ← R
4)      ∀x ∈ (ℂ − R)
5)          if γ_{R∪{x}}(𝔻) > γ_T(𝔻)
6)              T ← R ∪ {x}
7)      R ← T
8)  until γ_R(𝔻) == γ_ℂ(𝔻)
9)  return R
```

It starts off with an empty set and adds in turn, one at a time, those attributes that result in the greatest increase in the rough set dependency metric, until this produces its maximum possible value for the dataset. Other such techniques may be found in [18, 19].

2.7 Discernibility matrix

Many applications of rough sets make use of discernibility matrices for finding rules or reducts. A discernibility matrix[20, 21] of a decision table $(\mathbb{U}, \mathbb{C} \cup \mathbb{D})$ is a symmetric $|\mathbb{U}| \times |\mathbb{U}|$ matrix with entries defined by

$$c_{ij} = \{a \in \mathbb{C} | a(x_i) \neq a(x_j)\}, \quad i,j = 1,...,|\mathbb{U}|. \tag{13}$$

Each c_{ij} contains those attributes that differ between objects i and j.

For finding reducts, the so-called decision-relative discernibility matrix is of more interest. This only considers those object discernibility that occur when the corresponding decision attributes differ. Returning to the example dataset, the decision-relative discernibility matrix is produced, found in Table 2. For example, it can be seen from the table that objects 0 and 1 differ in each attribute. Although some attributes in objects 1 and 3 differ, their corresponding decisions are the same so no entry appears in the decision-relative matrix. Grouping all entries containing single attributes forms the core of the dataset (those attributes appearing in *every* reduct). Here, the core of the dataset is $\{d\}$.

Table 2 The decision-relative discernibility matrix

$x \in \mathbb{U}$	0	1	2	3	4	5	6	7
0								
1	a, b, c, d							
2	a, c, d	a, b, c						
3	b, c		a, b, d					
4	d	a, b, c, d		b, c, d				
5	a, b, c, d	a, b, c		a, b, d				
6	a, b, c, d		b, c		a, b, c, d	b, c		
7	a, b, c, d	d		a, c, d			a, d	

From this, the concept of discernibility functions can be introduced. This is a concise notation of how each object within the dataset may be distinguished from the others. A discernibility function f_D is a boolean function of m boolean variables $a_1^*, ..., a_m^*$ (corresponding to the membership of attributes $a_1, ..., a_m$ to a given entry of the discernibility matrix) defined as below:

$$f_D(a_1^*, ..., a_m^*) = \wedge\{\vee c_{ij}^* | 1 \leq j \leq i \leq |\mathbb{U}|, \ \ c_{ij} \neq \emptyset\} \quad (14)$$

where $c_{ij}^* = \{a^* | a \in c_{ij}\}$. By finding the set of all prime implicants of the discernibility function, all the minimal reducts of a system may be determined. From Table 2, the decision-relative discernibility function is (with duplicates removed)

$$\begin{aligned} f_D(a^*, b^*, c^*, d^*) = {} & (a^* \vee b^* \vee c^* \vee d^*) \wedge (a^* \vee c^* \vee d^*) \wedge \\ & (b^* \vee c^*) \wedge (d^*) \wedge (a^* \vee b^* \vee c^*) \wedge \\ & (a^* \vee b^* \vee d^*) \wedge (b^* \vee c^* \vee d^*) \wedge (a^* \vee d^*). \end{aligned}$$

Further simplification can be performed by removing those clauses that are subsumed by others:

$$f_D(a^*, b^*, c^*, d^*) = (b^* \vee c^*) \wedge (d^*).$$

The reducts of the dataset may be obtained by converting the above expression from conjunctive normal form to disjunctive normal form (without negations). Hence, the minimal reducts are $\{b, d\}$ and $\{c, d\}$. Although this is guaranteed to discover all minimal subsets, it is a costly operation rendering the method impractical for even medium-sized datasets. As all that is required is the discovery of a single reduct for many applications, efficient heuristic methods may be applied.

3 Rough set extensions

The reliance on discrete data for the successful operation of RST can be seen as a significant drawback of the approach. Indeed, this requirement of RST implies an objectivity in the data that is simply not present[22]. For example, in a medical dataset, values such as *Yes* or *No* cannot be considered objective for a Headache attribute as it may not be straightforward to decide whether a person has a headache or not to a high degree of accuracy. Again, consider an attribute Blood Pressure. In the real world, this is a real-valued measurement but for the purposes of RST must be discretised into a small set of labels such as "Normal", "High", *etc.* Subjective judgments are required for establishing boundaries for objective measurements.

In the rough set literature, several extensions have been developed that attempt to handle better the uncertainty present in real world data. In particular, variable precision rough sets[23] is a generalized model of rough sets, allowing a controlled degree of misclassification by relaxing the subset operator. Fuzzy-rough sets[24] and tolerance rough sets[25] handle real-valued data by replacing the traditional equivalence classes of crisp rough set theory with alternatives that are better suited to dealing with this type of data. In the fuzzy-rough case, fuzzy equivalence classes are employed within a fuzzy extension of rough set theory, resulting in a hybrid approach. In the tolerance case, indiscernibility relations are replaced with similarity relations that permit a limited degree of variability in attribute values. Approximations are constructed based on these tolerance classes in a manner similar to that of traditional rough set theory. An overview of these extensions is given below.

3.1 Variable precision rough sets

Variable precision rough sets (VPRS)[23] attempts to improve upon rough set theory by relaxing the subset operator. It was proposed to analyse and identify data patterns which represent statistical trends rather than functional. The main idea of VPRS is to allow objects to be classified with an error smaller than a certain predefined level.

This approach is arguably easiest to be understood within the framework of classification. Let $X, Y \subseteq \mathbb{U}$, the relative classification error is defined by

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

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

$$X \subseteq_\beta Y \ \text{ iff } \ c(X, Y) \leq \beta, \quad 0 \leq \beta < 0.5.$$

Using $\subseteq_\beta$ instead of $\subseteq$, the β-upper and β-lower approximations of a set X can be defined as

$$\begin{aligned} \underline{R}_\beta X &= \bigcup \{[x]_R \in \mathbb{U}/R | [x]_R \subseteq_\beta X\} \\ \overline{R}_\beta X &= \bigcup \{[x]_R \in \mathbb{U}/R | c([x]_R, X) < 1 - \beta\}. \end{aligned}$$

Note that $\underline{R}_\beta X = \underline{R} X$ for $\beta = 0$. The positive, negative and boundary regions in the original rough set theory can

now be extended to

$$POS_{R,\beta}(X) = \underline{R}_\beta X \quad (15)$$

$$NEG_{R,\beta}(X) = \mathbb{U} - \overline{R}_\beta X \quad (16)$$

$$BND_{R,\beta}(X) = \overline{R}_\beta X - \underline{R}_\beta X. \quad (17)$$

Returning to the example dataset in Table 1, (15) can be used to calculate the β-positive region for $R = \{b, c\}$, $X = \{e\}$ and $\beta = 0.4$. Setting β 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

$$\mathbb{U}/R = \{\{2\}, \{0, 4\}, \{3\}, \{1, 6, 7\}, \{5\}\}$$
$$\mathbb{U}/X = \{\{0\}, \{1, 3, 6\}, \{2, 4, 5, 7\}\}.$$

For each set $A \in \mathbb{U}/R$ and $B \in \mathbb{U}/X$, the value of c(A,B) must be less than β if the equivalence class A is to be included in the β-positive region. Considering $A = \{2\}$ gives

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

So object 2 is added to the β-positive region as it is a β-subset of $\{2, 4, 5, 7\}$ (and is in fact a traditional subset of the equivalence class). Taking $A = \{1, 6, 7\}$, a more interesting case is encountered:

$$c(\{1, 6, 7\}, \{0\}) = 1 > \beta$$
$$c(\{1, 6, 7\}, \{1, 3, 6\}) = 0.3333 < \beta$$
$$c(\{1, 6, 7\}, \{2, 4, 5, 7\}) = 0.6667 > \beta.$$

Here the objects 1, 6 and 7 are included in the β-positive region as the set $\{1, 6, 7\}$ is a β-subset of $\{1, 3, 6\}$. Calculating the subsets in this way leads to the following β-positive region:

$$POS_{R,\beta}(X) = \{1, 2, 3, 5, 6, 7\}.$$

Compare this with the positive region generated previously: $\{2, 3, 5\}$. Objects 1, 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 β-positive region of an equivalence relation Q on $\mathbb{U}$ may be determined by

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

where R is also an equivalence relation on $\mathbb{U}$. This can then be used to calculate dependencies and thus determine β-reducts. The dependency function becomes

$$\gamma_{R,\beta}(Q) = \frac{|POS_{R,\beta}(Q)|}{|\mathbb{U}|}.$$

It can be seen that the QuickReduct algorithm outlined previously can be adapted to incorporate the reduction method built upon VPRS theory. By supplying a suitable β value to the algorithm, the β-lower approximation, β-positive region, and β-dependency can replace the traditional calculations. This will result in a more approximate final reduct, which may be a better generalization when encountering unseen data. Additionally, setting β to 0 forces such a method to behave exactly like standard rough set theory.

Extended classification of reducts in the VPRS approach may be found in [26-28]. However, the variable precision approach requires the additional parameter β which has to be specified from the start. By repeated experimentation, this parameter can be suitably approximated. Nevertheless, problems arise when searching for true reducts as VPRS incorporates an element of imprecision in determining the number of classifiable objects.

3.2 Tolerance rough sets

Another way of attempting to handle imprecision is to introduce a measure of similarity of attribute values and define the lower and upper approximations based on these similarity measures.

3.2.1 Similarity measures

In this approach, suitable similarity relations must be defined for each attribute, although the same definition can be used for all attributes if applicable. A standard measure for this purpose, given in [29] is

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

where a is the attribute under consideration, and $a_{\max}$ and $a_{\min}$ denote the maximum and minimum values respectively for this attribute.

When considering more than one attribute, the defined similarities must be combined to provide a measure of the overall similarity of objects. For a subset of attributes, P, this can be achieved in many ways; two commonly adopted approaches are

$$(x, y) \in SIM_{P,\tau} \quad \text{iff} \quad \prod_{a \in P} SIM_a(x, y) \geq \tau \quad (19)$$

$$(x, y) \in SIM_{P,\tau} \quad \text{iff} \quad \frac{\sum_{a \in P} SIM_a(x, y)}{|P|} \geq \tau \quad (20)$$

where τ is a global similarity threshold. This framework allows for the specific case of traditional rough sets by defining a suitable similarity measure (e.g. equality of attribute values and (19)) and threshold ($\tau = 1$). Further similarity relations are investigated in [30], but are omitted here.

From this, the so-called tolerance classes that are generated by a given similarity relation for an object x are defined as

$$SIM_{P,\tau}(x) = \{y \in \mathbb{U} | (x, y) \in SIM_{P,\tau}\}. \quad (21)$$

3.2.2 Approximations and dependency

Lower and upper approximations are then defined in a similar way to traditional rough set theory:

$$\underline{P_\tau} X = \{x | SIM_{P,\tau}(x) \subseteq X\} \quad (22)$$

$$\overline{P_\tau} X = \{x | SIM_{P,\tau}(x) \cap X \neq \emptyset\}. \quad (23)$$

The tuple $\langle \underline{P_\tau} X, \overline{P_\tau} X \rangle$ is called a tolerance rough set[25]. Positive region and dependency functions then become

$$POS_{P,\tau}(Q) = \bigcup_{X \in \mathbb{U}/Q} \underline{P_\tau} X \tag{24}$$

$$\gamma_{P,\tau}(Q) = \frac{|POS_{P,\tau}(Q)|}{|\mathbb{U}|}. \tag{25}$$

From these definitions, methods for reduct search can be constructed that use the tolerance-based degree of dependency, $\gamma_{P,\tau}(Q)$, to gauge the significance of attribute subsets (in a similar way as QUICKREDUCT).

3.3 Fuzzy-rough sets

There have been two main lines of thought in the hybridization of fuzzy and rough sets, the constructive approach and the axiomatic approach. A general framework for the study of fuzzy-rough sets from both of these viewpoints is presented in [31]. For the constructive approach, generalized lower and upper approximations are defined based on fuzzy relations. Initially, these were fuzzy similarity/equivalence relations[24] but have since been extended to arbitrary fuzzy relations[31]. The axiomatic approach is primarily for the study of the mathematical properties of fuzzy-rough sets[32]. Here, various classes of fuzzy-rough approximation operators are characterized by different sets of axioms that guarantee the existence of types of fuzzy relations producing the same operators.

3.3.1 Main approaches

In the same way that crisp equivalence classes are central to rough sets, fuzzy equivalence classes are central to the fuzzy-rough set approach[24]. In classification applications for example, this means that the decision values and the conditional values may all be fuzzy. The concept of crisp equivalence classes can be extended by the inclusion of a fuzzy similarity relation S on the universe, which determines the extent to which two elements are similar in S[33]. The usual properties of reflexivity ($\mu_S(x,x) = 1$), symmetry ($\mu_S(x,y) = \mu_S(y,x)$) and transitivity ($\mu_S(x,z) \geq \mu_S(x,y) \wedge \mu_S(y,z)$, where $\wedge$ is a t-norm) hold.

Using such a fuzzy similarity relation S, the fuzzy equivalence class $[x]_S$ for objects close to x can be defined:

$$\mu_{[x]_S}(y) = \mu_S(x,y). \tag{26}$$

The following axioms should hold for a fuzzy equivalence class $F = [x]_S$[33]:

- $\exists x,\ \mu_F(x) = 1$
- $\mu_F(x) \wedge \mu_S(x,y) \leq \mu_F(y)$
- $\mu_F(x) \wedge \mu_F(y) \leq \mu_S(x,y)$

The first axiom corresponds to the requirement that an equivalence class is non-empty. The second axiom states that elements in y's neighborhood are in the equivalence class of y. The final axiom states that any two elements in F are related *via* S. Obviously, this definition degenerates to the normal definition of equivalence classes when S is non-fuzzy.

An initial definition of fuzzy P-lower and P-upper approximations was given in [24] as follows:

$$\mu_{\underline{P}X}(F_i) = \inf_x \max\{1 - \mu_{F_i}(x), \mu_X(x)\} \quad \forall i \tag{27}$$

$$\mu_{\overline{P}X}(F_i) = \sup_x \min\{\mu_{F_i}(x), \mu_X(x)\} \quad \forall i \tag{28}$$

where F_i is a fuzzy equivalence class and X is the (fuzzy) concept to be approximated. The tuple $\langle \underline{P}X, \overline{P}X \rangle$ is called a fuzzy-rough set.

Also defined in the literature are rough-fuzzy sets[34], which can be seen to be a particular case of fuzzy-rough sets. A rough-fuzzy set is a generalization of a rough set, derived from the approximation of a fuzzy set in a crisp approximation space. In [35] it is argued that, to be consistent, the approximation of a crisp set in a fuzzy approximation space should be called a fuzzy-rough set, and the approximation of a fuzzy set in a crisp approximation space should be called a rough-fuzzy set, making the two models complementary. In this framework, the approximation of a fuzzy set in a fuzzy approximation space is considered to be a more general model, unifying the two theories. However, most researchers consider the traditional definition of fuzzy-rough sets in [24] as standard.

The specific use of min and max operators in the definitions above is expanded in [36], where a broad family of fuzzy-rough sets is constructed, with each member represented by a particular implicator and t-norm. The properties of three well-known implicators (S-, R- and QL-implicators) are investigated. For example, a fuzzy-rough lower approximation defined using the Łukasiewicz implicator (which is both an S- and R-implicator) is as follows:

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

Further investigations in this area can be found in [31, 37-39].

In [40,41], an axiomatic approach is taken, but restricted to fuzzy T-similarity relations (and hence fuzzy T-rough sets), where T is a lower semi-continuous triangular norm. The work of [42] investigates the properties of generalized fuzzy-rough sets, defining a pair of dual generalized fuzzy approximation operators based on arbitrary fuzzy relations. The approach presented in [43] 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 (for example, fuzzy equivalence, fuzzy similarity, *etc.*) produce different classes of fuzzy-rough set algebras.

The work in [44] generalizes the fuzzy-rough set concept through the use of residual lattices. An arbitrary residual lattice L is used as a basic algebraic structure, and several classes of L-fuzzy-rough sets, defined using the product operator and its residuum provided by the residual lattice, and their properties are investigated. In [45], 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.

The relationships between fuzzy-rough set models and fuzzy ([0,1]-) topologies on a finite universe have been investigated. The first such research was reported in [40], 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 [31] investigated this for arbitrary fuzzy relations. In [46, 47] 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.

3.3.2 Other generalizations

In addition to the previous approaches to fuzzy-rough or rough-fuzzy hybridization, other generalizations are possible. One of the first attempts at hybridizing the two theories is reported in [10], 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. Those that are contained in the negative region (and therefore do not belong to the rough set) have zero membership. In so doing, a rough set can be expressed as a fuzzy set, with suitable modifications to the rough union and intersection operators. In [48], a definition of fuzzy-rough sets is given based on an algebraic approach to rough sets[49], where a rough set is defined as a pair of subsets from a sub-Boolean algebra without reference to the universe. The lower and upper bounds of such a rough set are then fuzzified. As stated in [35], the precise meaning of the upper and lower bounds may not be clear.

Another approach that blurs the distinction between rough and fuzzy sets has been proposed in [50]. The research was fueled by the concern that a purely numeric fuzzy set representation may be too precise, a concept is described exactly once its membership function has been defined. It seems as though excessive precision is required in order to describe imprecise concepts. The solution proposed is termed a shadowed set, which itself does not use exact membership values but instead employs basic truth values and a zone of uncertainty (the unit interval). A shadowed set could be thought of as an approximation of a fuzzy set or family of fuzzy sets where elements may belong to the set with certainty (membership of 1), possibility (unit interval) or not at all (membership of 0). This can be seen to be analogous to the definitions of the rough set regions: the positive region (certainty), the boundary region (possibility) and the negative region (no membership).

4 Applications

This section provides a brief overview of some of the many applications of rough set theory. There are several properties of rough sets that make the theory an obvious choice for use in dealing with real problems; for example, it handles uncertainty present in real data through approximations and also does not require threshold information in order to operate (as is the case with many current techniques).

4.1 Prediction of business failure

Attempts to develop business failure prediction models began seriously sometime in the late 1960s and continue through today. Although there has been much research in this area, there is still no unified well-specified theory of how and why corporations fail. Financial organizations need these predictions for evaluating firms of interest.

Many methods have been used for the purpose of bankruptcy prediction, such as logic analysis, discriminant analysis and probit analysis[51]. A comprehensive review of the various approaches to modeling and predicting this is presented in [52]. Although some of these methods led to satisfactory models, they suffered from limitations, often due to unrealistic statistical assumptions. Because of this, the rough set model, with its aim of keeping model assumptions to a minimum, appeared to be a highly useful approach for the analysis of financial information tables.

Rough set-based failure prediction was investigated in [53, 54]. In these investigations, the rough set approach was evaluated against several other methods, including C4.5[55], discriminant analysis and logit analysis. For the rough approach, decision rules were generated from the reducts produced by analysis of the financial information. All methods were then evaluated on data from the previous three years. The rough set model was found to be more accurate than discriminant analysis by an average of 6.1% per case, using a minimal set of reduced rules. It also outperformed C4.5, but performed similarly to logit analysis.

A comparative study of the rough sets model versus multi-variable discriminant analysis (MDA) can be found in [56]. It was demonstrated that through the use of rough set theory, the prediction of corporate bankruptcy was 97.0% accurate - an improvement over MDA which achieved an accuracy of 96.0%.

4.2 Financial investment

Trading systems have been built using rough set approaches. In [57-59], the rough set model was applied to discover strong trading rules that reflect highly repetitive patterns in data. Historical data from the Toronto stock exchange in 1980 was used for the extraction of trading rules for five companies. Experts confirmed that the extracted rules described the stock behavior and market sensitivity of these companies. Depending on a roughness parameter, the rules generated were either “general” or “exact”. The general rules were all recognised relationships in the investment industry, whereas the exact rules made less sense.

In the work reported in [60], the problem of how to deduce rules that map the financial indicators at the end of a month to the stock price changes a month later was addressed. This was based on 15 market indicators. From this study, only a satisfactory performance was achieved with many issues still to be tackled, such as data filtration and how to handle missing data. In [61], research was carried out into rough set reduct analysis and rule construction for forecasting the total index of the Oslo stock exchange. This also achieved satisfactory results, with a highest accuracy of 45%.

Research has been carried out on building trading systems for the S&P index[62]. Here, a hybrid system was developed that incorporated both neural networks and rough sets. Rules generated by rough sets were used to supervise neural networks to correct for possible errors in predictions. This system reduced drawdown by 25-50% and increased

the average winner/loser ratio by 50-100%.

Rough sets have also been applied to financial decision analysis and explanation for an industrial development bank, ETEVA[63, 64]. The bank was interested in investing its capital in firms, whilst reducing the risk involved in such an investment. To achieve this, a rough set-based firm assessment system was constructed that decided, based on a number of financial ratios, whether a company was acceptable, unacceptable or uncertain. An information table was constructed with the help of the financial manager of ETEVA. From this, the rough set-generated rules revealed the financial policy applied in the selection of firms. The rules can also be used to evaluate new firms that seek financing from the bank.

4.3 Bioinformatics and medicine

A common and diagnostically challenging problem facing emergency department personnel in hospitals is that of acute abdominal pain in children. There are many potential causes for this pain - most are usually non-serious. However, the pain may be an indicator that a patient has a serious illness, requiring immediate treatment and possibly surgery. Experienced doctors will use a variety of relevant historical information and physical observations to assess children. Such attributes occur frequently in recognizable patterns, allowing a quick and efficient diagnosis. Inexperienced physicians, on the other hand, may lack the knowledge and information to be able to recognize these patterns. The techniques developed in [65] provide a rough set-based clinical decision model to assist such inexperienced physicians. In this research, rough sets are used to support diagnosis by distinguishing between three disposition categories: discharge, observation/further investigation, and consult. Preliminary results show that the system gives an accuracy comparable to physicians, though it is dependent on a suitably high data quality.

Rough set data analysis is also applied to the problem of extracting protein-protein interaction sentences in biomedical literature[66]. Due to the abundance of published information relevant to this area, manual information extraction is a formidable task. This approach develops decision rules of protein names, interaction words, and their mutual positions in sentences. To increase the set of potential interaction words, a morphological model is developed, generating spelling and inflection variants. The performance of the method is evaluated using a hand-tagged dataset containing 1894 sentences, producing a precision-recall break-even performance of 79.8% with leave-one-out cross-validation.

Automated classification of calculated electroencephalogram (EEG) parameters has been shown to be a promising method for detection of intraoperative awareness. In [67], rough set-based methods were employed to generate classification rules resulting in satisfactory accuracy rates of approximately 90%.

Gene expression experiments, where the genetic content of samples is obtained with high throughput technologies, result in high dimensional data. For useful information to be discovered from this data (usually comprising of thousands of genes), automated methods must be able to either cope with this dimensionality or reduce it intelligently. Typically, the latter option is chosen as this has the additional benefit of making the extracted knowledge more readable. Many rough set-based methods have been applied to this task - both for feature reduction and classification rule discovery[68, 69].

4.4 Fault diagnosis

A rough set approach for the diagnosis of valve faults in a multi-cylinder diesel engine is investigated in [70]. The use of rough sets enabled the diagnosis of several fault categories in a generic manner. A decision table was constructed from attributes extracted from the vibration signals, with four operational states studied among the signal characteristics: normal, intake valve clearance too small, intake valve clearance too large, exhaust valve clearance too large. Three sampling points were selected for the collection of vibration signals. The results demonstrated that the system is quite effective for such fault diagnosis, and the extracted rules correspond well with prior knowledge of the system.

In [71], a rough set-based method for continuous failure diagnosis in assembly systems is presented. Sensor measurements were used to construct a diagnosis table from which rough set rules were extracted.

4.5 Spacial and meteorological pattern classification

Sunspot observation, analysis and classification form an important part in furthering knowledge about the sun, the solar weather, and its effect on earth. Certain categories of sunspot groups are associated with solar flares. Observatories around the world track all visible sunspots in an effort to early detect flares. Sunspot recognition and classification are currently manual and labor intensive processes which could be automated if successfully learned by a machine. The approach presented in [72] employs a hierarchical rough set-based learning method for sunspot classification. It attempts to learn the modified Zurich classification scheme through rough set-based decision tree induction. The resulting system is evaluated on sunspots extracted from satellite images, with promising results.

In [73], a new application of rough set theory for classifying meteorological radar data is introduced. Volumetric radar data is used to detect storm events responsible for severe weather. Classifying storm cells is a difficult problem as they exhibit a complex evolution throughout their lifespan. Also, the high dimensionality and imprecision of the data can be prohibitive. Here, a rough set approach is employed to classify a number of meteorological storm events.

4.6 Music and acoustics

A dominance-based rough set approach, an extension of rough sets to preference-ordered information systems, was used in [74] to generate preference models for violin quality grading. A set of violins were submitted to a violinmaker′s competition and evaluated by a jury according to several assessment criteria. The sound of the instruments was recorded digitally and then processed to obtain sound attributes. These features, along with jury assessments were analyzed by the rough set method, generating preference models. It was shown that the jury′s rankings were

well approximated by the automated approach.

In [75], an approach to classifying swallowing sound signals is given, utilising rough set theory. This approach has been developed to facilitate the detection of patients at risk of aspiration, or choking. The waveform dimension is used to describe sound signal complexity and major changes in signal variance. From swallow sound data tables, decision rules were derived, *via* rough sets. The algorithms yielded a high classification accuracy, whilst producing a comparatively small ruleset.

A decision system employing rough sets and neural networks is presented in [76]. The aim of the study was to automatically classify musical instrument sounds on the basis of a limited number of parameters, and to test the quality of musical sound parameters that are included in the MPEG-7 standard. The use of wavelet-based parameters led to better audio retrieval efficiency.

The classification of musical works is considered in [77], based on the inspection of standard music notations. A decision table is constructed, with features representing various aspects of musical compositions (objects), such as rhythm disorder, beat characteristics and harmony. From this, classification rules are induced (*via* rough set rule induction) and used to classify unseen compositions.

4.7 Feature selection

The discussion above has focussed on actual practical applications of rough set theory. This section is concerned with the theoretical advancement of feature selection within the rough set community.

As indicated previously, the work on rough set theory offers a formal methodology that can be employed to reduce the dimensionality of datasets, often as a preprocessing step to assist other tasks like learning from data[78]. The QuickReduct algorithm provided earlier is a typical example of rough set-assisted feature selection tools. Such a method helps select the most information rich features in a dataset, without transforming the data, all while attempting to minimize information loss during the selection process. Computationally, the approach is highly efficient as it involves simple set operations only. Thus, it represents one of the most successful applications of rough sets. However, it is reliant upon a discrete dataset; important information may be lost as a result of quantization of the underlying numerical features (that real-world problems typically have). It is natural, then, to apply its extensions to this area.

Such research has been carried out in [79-81], where a reduction method was proposed based on fuzzy extensions to the positive region and dependency function based on fuzzy lower approximations. A greedy hill-climber is used to perform subset search, using the fuzzy dependency function both for subset evaluation and as a stopping criterion. The method was used successfully within a range of problem domains, including web content classification and complex system monitoring[80].

Optimizations are given in [80,82] to improve the performance of the method. In [83], a compact computational domain is proposed to reduce the computational effort required to calculate fuzzy lower approximations for large datasets, based on some of the properties of fuzzy connectives. Fuzzy entropy is used in [84] to guide the search toward smaller reducts. In [82], an alternative search algorithm is presented that alleviates some of the problems encountered with a greedy hill-climber approach. This problem is also tackled in [85] *via* the use of a novel ant colony optimization-based framework for feature selection. A genetic algorithm is used in [86] for search based on the fuzzy dependency function within a face recognition system with promising results.

The work in [87,88] improves upon these developments by formally defining relative reductions for fuzzy decision systems. A discernibility matrix is constructed for the computation of all such reductions. As the resulting discernibility matrix is crisp, some information may have been lost in this process. Additionally, there are complexity issues when computing discernibility matrices for large datasets. However, in the crisp rough set literature there have been methods proposed that avoid this[30, 89], and similar constructions may be applicable here.

Feature selection algorithms, based on the generalization of fuzzy approximation spaces to fuzzy probability approximation spaces are introduced in [90]. This is achieved through the introduction of a probability distribution on the universe. Information measures for fuzzy indiscernibility relations are presented in [91] for the computation of feature importance. Reducts are computed through the use of a greedy selection algorithm similar to QuickReduct.

5 Conclusion

This paper has presented an overview of the rough set theory and its extensions, supported with a brief discussion of a number of representative applications of these theories. In particular, the paper has introduced the basic rough set concepts of indiscernibility; lower and upper approximations; positive, negative and boundary regions; attribute dependency and significance; reducts and discernibility matrix. These notions are useful to develop automated computational information and decision systems.

Because of the clear advantage of rough sets in performing data and information analysis without the need of preliminary information about data (e.g. probabilities in statistics, probabilistic assignments in Dempster-Shafer theory, and membership functions in fuzzy set theory), despite its recency, the seminal rough set theory has been extended in various ways to further its potential. This paper has given an outline of three such approaches, including variable precision rough sets, tolerance rough sets and fuzzy rough sets. These extensions allow the ability of the original rough set theory in handling discrete and nominal data, which is assumed to be true and accurate reflection of the world, to be maximized to cope with numerical and other contextual aspects of real world data.

To demonstrate the success of rough sets and their extensions in making use of imperfect knowledge to solve practical problems, the paper has also provided a short account of some representative applications. As an exciting developing discipline, there are nevertheless many areas in which much research may be carried out to improve further the mathematical rigorousness and the computational power associated with the many techniques derived from the original theory and its extensions. For example, fuzzification

of more rough set concepts should provide further flexible techniques, such as fuzzy discernibility matrices and functions. Indeed, by extending those concepts fundamental to crisp rough set rule construction, approaches to fuzzy-rough rule induction may be developed, offering more flexibility and comprehensibility. Additionally, there is great potential for developing cross-hybrid approaches - where two separate hybrid extensions to rough set theory are themselves hybridized. Preliminary work in this area has focussed on the hybridization of variable precision rough sets and fuzzy-rough sets [92]. Such research will no doubt help advance the applications of rough sets in even wide-reaching areas.

References

[1] Z. Pawlak. Rough Sets. *International Journal of Computer and Information Sciences*, no. 11, pp. 341–356, 1982.

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

[3] Z. Pawlak. Some Issues on Rough Sets. *Transactions on Rough Sets*, vol. 1, pp. 1–58, 2004.

[4] I. Düntsch, G. Gediga. Rough Set Data Analysis. *Encyclopedia of Computer Science and Technology*, vol. 43, no. 28, pp. 281–301, 2000.

[5] S. Hayri, V. V. Raghavan, T. D. Johnsten. The Status of Research on Rough Sets for Knowledge Discovery in Databases. In: *Proceedings of the Second International Conference on Nonlinear Problems in Aviation and Aerospace*, Daytona Beach, Florida, USA, vol. 2, pp. 673–680, 1998.

[6] A. Skowron, Z. Pawlak, J. Komorowski, L. Polkowski. A Rough Set Perspective on Data and Knowledge. *Handbook of Data Mining and Knowledge Discovery*, Oxford University Press, pp. 134–149, 2002.

[7] S. K. Pal, A. Skowron. *Pattern Recognition Letters-Special Issue on Rough Sets, Pattern Recognition and Data Mining*, vol. 24, no. 6, pp. 829–933, 2003.

[8] A. Skowron, J. W. Grzymala-Busse. From Rough Set Theory to Evidence Theory. In *Advances in the Dempster-Shafer Theory of Evidence*, R. Yager, M. Fedrizzi, and J. Kasprzyk (eds.), John Wiley & Sons Inc., New York, 1994.

[9] Z. Pawlak, A. Skowron. Rough Membership Functions. In *Advances in the Dempster-Shafer Theory of Evidence*, R. Yager, M. Fedrizzi, J. Kacprzyk (eds.), Wiley, New York, pp. 251–271, 1994.

[10] M. Wygralak. Rough Sets and Fuzzy Sets – Some Remarks on Interrelations. *Fuzzy Sets and Systems*, vol. 29, no. 2, pp. 241–243, 1989.

[11] A. P. Dempster. A Generalization of Bayesian Inference. *Journal of the Royal Statistical Society, Series B*, vol. 30, no. 2, pp. 205–247, 1968.

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

[13] L.A. Zadeh. Fuzzy Sets. *Information and Control*, vol. 8, no.3, pp. 338–353, 1965.

[14] M. Dash, H. Liu. Feature Selection for Classification. *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997.

[15] J. J. Alpigini, J. F. Peters, A. Skowron, N. Zhong, (eds). *Rough Sets and Current Trends in Computing*, Springer, 2002.

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

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

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

[19] L. Polkowski. *Rough Sets: Mathematical Foundations*, Physica-Verlag, Heidelberg, 2002.

[20] J. Komorowski, Z. Pawlak, L. Polkowski, A. Skowron. Rough Sets: A Tutorial. In *Rough-Fuzzy Hybridization: A New Trend in Decision Making*, Springer-Verlag Singapore Ltd, pp. 3–98, 1999.

[21] A. Skowron, C. Rauszer. The Discernibility Matrices and Functions in Information Systems. In *Intelligent Decision Support-handbook of Applications and Advances of the Rought Set Theory*, Kluwer Academic Publisher, Dordrecht, pp. 331–362, 1992.

[22] W. W. Koczkodaj, M. Orlowski, V. W. Marek. Myths about Rough Set Theory. *Communications of the ACM*, vol. 41, no. 11, pp. 102–103, 1998.

[23] W. Ziarko. Variable Precision Rough Set Model. *Journal of Computer and System Sciences*, vol. 46, no. 1, pp. 39–59, 1993.

[24] D. Dubois, H. Prade. Putting Rough Sets and Fuzzy Sets Together. *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, pp. 203–232, 1992.

[25] A. Skowron, J. Stepaniuk. Tolerance Approximation Spaces. *Fundamenta Informaticae*, vol. 27, no. 2–3, pp. 245–253, 1996.

[26] M. J. Beynon. An Investigation of β-reduct Selection within the Variable Precision Rough Sets Model. In *Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing*, pp 114–122, 2000.

[27] M. Beynon. Reducts within the Variable Precision Rough Sets Model: A Further Investigation. *European Journal of Operational Research*, vol. 134, no. 3, pp. 592–605, 2001.

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

[29] J. Stefanowski. On Rough Set Based Approaches to Induction of Decision Rules. In *Rough Sets in Knowledge Discovery*, vol. 1, Physica Verlag, Heidelberg, pp. 500–529, 1998.

[30] H. S. Nguyen, A. Skowron. Boolean Reasoning for Feature Extraction Problems. In *Proceedings of the 10th International Symposium on Foundations of Intelligent Systems*, Springer-Verlag, London, UK, pp. 117–126, 1997.

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

[32] W. Z. Wu, W. X. Zhang. Constructive and Axiomatic Approaches of Fuzzy Approximation Operators. *Information Sciences*, vol. 159, no.3-4, pp. 233–254, 2004.

[33] U. Höhle. Quotients with Respect to Similarity Relations. *Fuzzy Sets and Systems*, vol. 27, no. 1, pp. 31–44, 1988.

[34] P. Srinivasan, M. E. Ruiz, D. H. Kraft, J. Chen. Vocabulary Mining for Information Retrieval: Rough Sets and Fuzzy Sets. *Information Processing & Management*, vol. 37, no. 1, pp. 15–38, 1998.

[35] Y. Y. Yao. Combination of Rough and Fuzzy Sets Based on α-Level Sets. In *Rough Sets and Data Mining: Analysis of Imprecise Data*, T. Y. Lin, N. Cereone (eds.), Kluwer Academic Publishers, pp. 301–321, 1997.

[36] A. M. Radzikowska, E. E. Kerre. A Comparative Study of Fuzzy Rough Sets. *Fuzzy Sets and Systems*, vol. 126, no. 2, pp. 137–155, 2002.

[37] M. De Cock, C. Cornelis, E. E. Kerre. Fuzzy Rough Sets: Beyond the Obvious. In *IEEE International Conference on Fuzzy Systems*, vol. 1, pp. 103–108, 2004.

[38] H. Thiele. Fuzzy Rough Sets Versus Rough Fuzzy Sets - an Interpretation and a Comparative Study Using Concepts of Modal Logics. Technical report CI-30/98, University of Dortmund, 1998.

[39] W. Z. Wu, Y. Leung, J. S. Mi. On Characterizations of (I,T)-Fuzzy Rough Approximation Operators. *Fuzzy Sets and Systems*, vol. 154, no. 1, pp. 76–102, 2005.

[40] D. Boixader, J. Jacas, J. Recasens. Upper and Lower Approximations of Fuzzy Sets. *International Journal of General Systems*, vol. 29, no. 4, pp. 555–568, 2000.

[41] N. N. Morsi, M. M. Yakout. Axiomatics for Fuzzy Rough Sets. *Fuzzy Sets and Systems*, vol. 100, no. 1-3, pp. 327–342, 1998.

[42] W. Z. Wu, J. S. Mi, W. X. Zhang. Generalized Fuzzy Rough Sets. *Information Sciences*, vol. 151, pp. 263–282, 2003.

[43] J. S. Mi, W. X. Zhang. An Axiomatic Characterization of a Fuzzy Generalization of Rough Sets. *Information Sciences*, vol. 160, no. 1-4, pp. 235–249, 2004.

[44] A. M. Radzikowska, E. E. Kerre. Fuzzy Rough Sets Based on Residuated Lattices. *Lecture Notes in Computer Science*, Springer Berlin, Heidelberg, vol. 3135, pp. 278–296, 2004.

[45] D. Chen, W. X. Zhang, D. Yeung, E. C. C. Tsang. Rough Approximations on a Complete Completely Distributive Lattice with Applications to Generalized Rough Sets. *Information Sciences*, vol. 176, no. 13, pp. 1829–1848, 2006.

[46] K. Qin, Z. Pei. On the Topological Properties of Fuzzy Rough Sets. *Fuzzy Sets and Systems*, vol. 151, no. 3, pp. 601–613, 2005.

[47] W. Z. Wu. A Study on Relationship Between Fuzzy Rough Approximation Operators and Fuzzy Topological Spaces, *Lecture Notes in Artificial Intelligence*, Springer-Verlag Heidelberg, vol.3613, pp. 167–174, 2005.

[48] S. Nanda, S. Majumdar. Fuzzy Rough Sets. *Fuzzy Sets and Systems*, vol. 45, pp. 157–160, 1992.

[49] T. B. Iwiński. Algebraic Approach to Rough Sets. *Bulletin of the Polish Academy of Science and Mathematics* vol. 35, pp. 673–683, 1987.

[50] W. Pedrycz. Shadowed Sets: Bridging Fuzzy and Rough Sets. In *Rough Fuzzy Hybridization: A New Trend in Decision-making*, S. K. Pal, A. Skowron (eds.), Springer Verlag, Singapore, pp. 179–199, 1999.

[51] F. E. H. Tay, L. Shen. Economic and Financial Prediction Using Rough Sets Model. *European Journal of Operational Research*, vol. 141, no. 3, pp. 641–659, 2002.

[52] A. I. Dimitras, S. H. Zanakis, C. Zopounidis. A Survey of Business Failure with an Emphasis on Prediction Methods and Industrial Applications. *European Journal of Operational Research*, vol. 90, no. 3, pp. 487–513, 1996.

[53] R. Slowinski, C. Zopounidis, A. I. Dimitras, R. Susmaga. Rough Set Predictor of Business Failure. In *Soft Computing in Financial Engineering*, R. A. Ribeiro, H. J. Zimmermann, R. R. Yager, J. Kacprzyk (eds.), Physica-Verlag, Wurzburg, pp. 402–424, 1999.

[54] A. I. Dimitras, R. Slowinski, R. Susmaga, C. Zopounidis. Business Failure Prediction Using Rough Sets. *European Journal of Operational Research*, vol. 114, no. 2, pp. 263–280, 1999.

[55] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[56] A. Szladow, D. Mills. Tapping Financial Databases. *Business Credit*, vol. 95, no. 7, p. 8, 1993.

[57] R. Golan, D. Edwards. Temporal Rules Discovery Using Datalogic/R+ with Stock Market Data. In *Proceedings of the International Workshop on Rough Sets and Knowledge Discovery*, Springer-Verlag, pp. 74–81, 1993.

[58] R. Golan. Stock Market Analysis Utilizing Rough Set Theory. Ph.D. dissertation, Department of Computer Science, University of Regina, Canada, 1995.

[59] W. Ziarko, R. Golan, D. Edwards. An Application of Datalogic/R Knowledge Discovery Tool to Identify Strong Predictive Rules in Stock Market Data. In *Proceedings of AAAI Workshop on Knowledge Discovery in Databases*, Washington DC, pp. 89–101, 1993.

[60] J. Bazan, A. Skowron, P. Synak. Dynamic Reducts As a Tool for Extracting Laws From Decision Tables. In *Proceedings of the 8th Symposium on Methodologies for Intelligent Systems*, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, vol. 869, pp. 346–355, 1994.

[61] J. K. Baltzersen. An Attempt to Predict Stock Market Data: A Rough Sets Approach. Master dissertation, Knowledge Systems Group, Department of Computer Systems and Telematics, The Norwegian Institute of Technology, University of Trondheim, 1996.

[62] M. Ruggiero. Turning the Key. *Futures*, vol. 23, no. 14, pp. 38–40, 1994.

[63] R. Slowinski, C. Zopounidis. Rough Set Sorting of Firms According to Bankruptcy Risk. In *Applying Multiple Criteria Aid for Decision to Environmental Management*, Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 339–357, 1994.

[64] R. Slowinski, C. Zopounidis. Application of the Rough Set Approach to Evaluation of Bankruptcy Risk. *International Journal of Intelligent Systems in Accounting, Finance & Management*, vol. 4, no. 1, pp. 27–41, 1995.

[65] K. Farion, W. Michalowski, R. Slowinski, S. Wilk, S. Rubin. Rough Set Methodology in Clinical Practice: Controlled Hospital Trial of the MET System. In *Proceedings of the 2004 International Conference on Rough Sets and Current Trends in Computing*, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Heidelberg, vol. 3066, pp. 805–814, 2004.

[66] F. Ginter, T. Pahikkala, S. Pyysalo, J. Boberg, J. Järvinen, T. Salakoski. Extracting Protein-Protein Interaction Sentences by Applying Rough Set Data Analysis. In *Proceedings of the 2004 International Conference on Rough Sets and Current Trends in Computing*, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Heidelberg, vol. 3066, pp. 780–785, 2004.

[67] M. Ningler, G. Stockmanns, G. Schneider, O. Dressler, E. F. Kochs. Rough Set-based Classification of EEG-Signals to Detect Intraoperative Awareness: Comparison of Fuzzy and Crisp Discretization of Real Value Attributes. In *Proceedings of the 2004 International Conference on Rough Sets and Current Trends in Computing*, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Heidelberg, vol. 3066, pp. 825–834, 2004.

[68] T. Hvidsten, A. Lægreid, J. Komorowski. Learning Rule-based Models of Biological Process from Gene Expression Time Profiles Using Gene Ontology. *Bioinformatics*, vol. 19, no. 9, pp. 1116–1123, 2003.

[69] H. Midelfart, H. J. Komorowski, K. Nørsett, F. Yadetie, A. K. Sandvik, A. Lægreid. Learning Rough Set Classifiers from Gene Expressions and Clinical Data. *Fundamenta Informaticae*, vol. 53, no. 2, pp. 155–183, 2002.

[70] L. Shen, F. E. H. Tay, L. Qu, Y. Shen. Fault Diagnosis Using Rough Sets Theory. *Computers in Industry*, vol. 43, no. 1, pp. 61–72, 2000.

[71] K. Mannar and D. Ceglarek. Continuous Failure Diagnosis for Assembly Systems Using Rough Set Approach. *Annals of the CIRP*, vol. 53, no. 1, pp. 39–42, 2004.

[72] S. H. Nguyen, T. T. Nguyen, H. S. Nguyen. Rough Set Approach to Sunspot Classification Problem. In *Proceedings of the 2005 International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Copmputing*, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Heidelberg, vol. 3642, pp. 263–272, 2005.

[73] J. F. Peters, Z. Suraj, S. Shan, S. Ramanna, W. Pedrycz, N. Pizzi. Classification of Meteorological Volumetric Radar Data Using Rough Set Methods. *Pattern Recognition Letters*, vol. 24, no. 6, pp. 911–920, 2003.

[74] J. Jelonek, E. Lukasik, A. Naganowski, R. Slowinski. Inducing Jury′s Preferences in Terms of Acoustic Features of Violin Sounds. In *Proceedings of the 2004 International Conference on Artificial Intelligence and Soft Computing*, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Heidelberg, vol. 3070, pp. 492–497, 2004.

[75] L. Lazareck, S. Ramanna. Classification of Swallowing Sound Signals: A Rough Set Approach. In *Proceedings of the 2004 International Conference on Rough Sets and Current Trends in Computing*, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Heidelberg, vol. 3066, pp. 679–684, 2004.

[76] B. Kostek, P. Szczuko, P. Zwan. Processing of Musical Data Employing Rough Sets and Artificial Neural Networks. In *Proceedings of the 2004 International Conference on Rough Sets and Current Trends in Computing*, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Heidelberg, vol. 3066, pp. 539–548, 2004.

[77] M. P. Hippe. Towards the Classification of Musical Works: A Rough Set Approach. In *Proceedings of the 2002 International Conference on Rough Sets and Current Trends in Computing*, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Heidelberg, vol. 2475, pp. 546–553, 2002.

[78] Q. Shen. Semantics-Preserving Dimensionality Reduction in Intelligent Modelling. In *Modelling with Words*, J. Lawry, J. Shanahan, A. Ralescu (eds.), Springer, pp. 64–79, 2003.

[79] R. Jensen, Q. Shen. Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough Based Approaches. *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 12, pp. 1457–1471, 2004.

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

[81] Q. Shen, R. Jensen. Selecting Informative Features with Fuzzy-Rough Sets and Its Application for Complex Systems Monitoring. *Pattern Recognition*, vol. 37, no. 7, pp. 1351–1363, 2004.

[82] R. B. Bhatt, M. Gopal. On Fuzzy-Rough Sets Approach to Feature Selection. *Pattern Recognition Letters*, vol. 26, no. 7, pp. 965–975, 2005.

[83] R. B. Bhatt, M. Gopal. On the Compact Computational Domain of Fuzzy-Rough Sets. *Pattern Recognition Letters*, vol. 26, no. 11, pp. 1632–1640, 2005.

[84] N. M. Parthalain, R. Jensen, Q. Shen. Fuzzy Entropy-assisted Fuzzy-Rough Feature Selection. In *Proceedings of the 2006 IEEE International Conference on Fuzzy Systems*, pp. 423–430, 2006.

[85] R. Jensen, Q. Shen. Fuzzy-Rough Data Reduction with Ant Colony Optimization. *Fuzzy Sets and Systems*, vol. 149, no. 1, pp. 5–20, 2005.

[86] L. Zhou, W. Li, Y. Wu. Face Recognition Based on Fuzzy Rough Set Reduction. In *Proceedings of the 2006 International Conference on Hybrid Information Technology*, vol. 1, pp. 642–646, 2006.

[87] G. C. Y. Tsang, D. Chen, E. C. C. Tsang, J. W. T. Lee, D. S. Yeung. On Attributes Reduction with Fuzzy Rough Sets. In *Proceedings of the 2005 IEEE International Conference on Systems, Man, and Cybernetics*, vol. 3, pp. 2775–2780, 2005.

[88] X. Z. Wang, Y. Ha, D. Chen. On the Reduction of Fuzzy Rough Sets. In *Proceedings of the 2005 International Conference on Machine Learning and Cybernetics*, vol. 5, pp. 3174–3178, 2005.

[89] Jue. Wang, Ju. Wang. Reduction Algorithms Based on Discernibility Matrix: The Ordered Attributes Method. *Journal of Computer Science & Technology*, vol. 16, no. 6, pp. 489–504, 2001.

[90] Q. Hu, D. Yu, Z. Xie, J. Liu. Fuzzy Probabilistic Approximation Spaces and Their Information Measures. *IEEE Transactions on Fuzzy Systems*, vol. 14, no.2, pp. 191–201, 2006.

[91] Q. Hu, D. Yu, Z. Xie. Information-preserving Hybrid Data Reduction Based on Fuzzy-Rough Techniques. *Pattern Recognition Letters*, vol. 27, no. 5, pp. 414–423, 2006.

[92] A. Mieszkowicz-Rolka, L. Rolka. Variable Precision Fuzzy Rough Sets Model in the Analysis of Process Data. *Lecture Notes in Computer Science*, Springer, Heidelberg, vol. 3641, pp. 354–363, 2005.

Qiang Shen received the B.Sc. and M.Sc. degrees in communications and electronic engineering from the National University of Defence Technology, China, and the Ph.D. degree in knowledge-based systems from Heriot-Watt University, Edinburgh, U.K.

He is a professor with the Department of Computer Science at the University of Wales, Aberystwyth, and an honorary fellow at the University of Edinburgh. His research interests include fuzzy and imprecise modeling, model-based inference, pattern recognition, and knowledge refinement and reuse. Professor Shen is an associate editor of the *IEEE Transactions on Fuzzy Systems* and of the *IEEE Transactions on Systems, Man, and Cybernetics (Part B)*, and an editorial board member of the *Fuzzy Sets and Systems* Journal amongst others. He has published over 180 papers in academic journals and conferences on topics within artificial intelligence and related areas.

Richard Jensen received the B.Sc. degree in computer science from Lancaster University, U.K., and the M.Sc. and Ph.D. degrees in artificial intelligence from the University of Edinburgh, U.K.

He is a research fellow with the Department of Computer Science at the University of Wales, Aberystwyth, working in the Advanced Reasoning Group. His research interests include rough and fuzzy set theory, pattern recognition, information retrieval, feature selection, and swarm intelligence. He has published over 20 papers in these areas.