JOURNAL OF COMPUTER AND SYSTEM SCIENCES 46, 39–59 (1993)

Variable Precision Rough Set Model

WOJCIECH ZIARKO

Computer Science Department, University of Regina, Regina, Saskatchewan, Canada

Received June 1, 1990; revised August 1, 1991

A generalized model of rough sets called variable precision model (VP-model), aimed at modelling classification problems involving uncertain or imprecise information, is presented. The generalized model inherits all basic mathematical properties of the original model introduced by Pawlak. The main concepts are introduced formally and illustrated with simple examples. The application of the model to analysis of knowledge representation systems is also discussed. © 1993 Academic Press, Inc.

1. INTRODUCTION

The theory of rough sets, as proposed by Pawlak [1], provides a formal tool for dealing with imprecise or incomplete information in terms of three valued logic. Since its introduction the theory has generated a great deal of interest along logicians [2, 3, 9] as well as among researchers dealing with machine learning and knowledge acquisition for expert systems [4–14, 18–21]. Substantial progress has been achieved in understanding practical implications and limitations of this approach. In particular, the inability to model uncertain information was one limitation frequently emphasized by users of the software package DataQuest. It is used for knowledge acquisition and analysis [12, 15] based on the theory of rough sets. This limitation severely reduces the applicability of the rough set approach to problems which are more probabilistic than deterministic in nature. An attempt to overcome this restriction was reported in [16]. However, the proposed generalization was based on strong statistical assumptions and did not directly inherit all useful properties of the original model of rough sets.

In this paper a new generalization of the rough set model is proposed. This generalization is aimed at handling undertain information and is directly derived from the original model without any additional assumptions. The properties of the new model are investigated, illustrated with examples, and related to the properties of the classical approach. The application areas for this model fall into the category of broadly understood knowledge discovery in databases, for example, for the purpose of rule induction from data, or control algorithm acquisition from analysis of previous operators' actions and pattern recognition.

0022-0000/93 $5.00

2. Motivation

2.1. *Some Limitations of the Rough Sets Model*

The central problem of the theory of rough sets is classification analysis. The whole approach is inspired by the notion of inadequacy of available information to perform complete classification of objects belonging to a specified category such as cars, humans, etc.

Quite frequently, the available information allows only for partial classification. The theory of rough sets can be used to model this kind of classification but the classification must be fully correct or certain. The classification with a controlled degree of uncertainty, or a misclassification error, is outside the realm of this approach. It seems, however, that admitting in practice some level of uncertainty in the classification process may lead to a deeper understanding and a better utilization of the properties of data being analysed. For example, if 90% of German-made cars are in a category of high quality cars then the classification rule associating the feature "German-made" with the category "high quality" should be given a high degree of confidence. In other words, when predicting "high quality," knowing that a car is "German-made," the likelihood of incorrect decision will be low. A similar idea of decision making with a predefined degree of uncertainty has been used in statistical theory to estimate unknown value of a parameter of a probability distribution [23]. The notion of confidence interval is applied in this case to specify a range of values which will contain, with probability above a predefined confidence level, a real value of an estimated distribution parameter. The confidence level, which is usually close to one, ensures that regardless of whichever value, belonging to the confidence interval, will be used as an estimation of the actual value of the parameter, the likelihood of an error; i.e., the actual value, being further than the interval radius from the estimated value, will be very low.

Another limitation of the original rough sets model stems from the assumption that the universe U of data objects under consideration is known and that all conclusions derived from the model are applicable only to this set of objects. In practice, however, there is an evident need to generalize conclusions obtained from a smaller set of example objects to a larger population. For example, the relationship between symptoms and diseases discovered from analysis of a number of past cases in a medical application may be used to draw a generalized conclusion which could, in turn, be applicable to new cases. This kind of induction is present in statistical reasoning and cannot be avoided here if the model is to be applicable to real life situations. Therefore, all conclusions derived from sample data are true only with respect to that set of data, and, they should be treated as uncertain hypotheses about properties of a larger universe.

Hypotheses derived from sample data should not, however, be based only on error-free classification rules observed in the sample data. Also, partially incorrect classification should be taken into account. Any partially incorrect classification rule provides valuable trend information if the majority of available data to which

such a rule applies can be correctly classified. This trend information can be captured and analysed by using a collection of techniques derived from the proposed extended model of rough sets referred to as a *variable precision rough set model* (VP-model).

The main objective of this article is to introduce the VP-model, to investigate some of the properties of the extended model, and to demonstrate how it can be used as a tool for data analysis. The generalized model allows for a controlled degree of misclassification in its formalism which, in turn, leads to more general notions of set approximations. The standard model of rough sets becomes a special case of VP-model. The primary advantage of a VP-model, in the context of data analysis applications, is its ability to recognize the presence of data dependencies in situations where data items are considered independent by the original rough sets model. Such situations occur when data dependencies are non-functional. Characterizing such non-functional, or non-deterministic dependencies in terms of approximate decision rules is one of the main practical reasons behind introduction of the VP-model of rough sets.

2.2. *Relationship to Other Classification Models*

The end result of the use of the rough sets model, or the VP-model, to data analysis is a set of classification rules for classifying objects into two or more categories. The rules form a description of each category, typically in terms of a Boolean formula combining simple predicates expressing some observable properties of objects, such as, the formula NR-OF-CYLINDERS > 2 AND ENGINE-POWER = HIGH, which is the description of a class of cars. New cases can be classified by matching their features with rule conditions. The main issue in the rough set approach is the formation of "good" rules, that is such rules which:

(1) discriminate between different decision categories (or partially discriminate between categories, such as, in a VP-model);

(2) capture essential factors affecting the classification result and do not take into account irrelevant factors;

(3) are nonredundant in terms of minimizing the required number of rules and their conditions;

(4) receive strong support in the available data by being matched by many "training" cases, i.e., are general;

(5) exhibit low error rate on new cases as a result of their generality.

The above set of objectives has some overlap with the objectives of the statistical approaches to the classification problem. However, the formation of characteristic and discriminating descriptions of a decision category is not an issue in statistical techniques. The main issue in the statistical approach is the construction of a probabilistic classifier approximating the theoretically optimal Bayes classification rule [23–26]. Different methods are used to build such classifiers [23–26]. They usually involve strong assumptions, such as the typical assumptions of Gaussian

distribution of feature values in parametric techniques [24], or require very large sample collections, such as the method of estimation of probability density in a non-parametric approaches. Some non-parametric techniques make a highly questionable assumption of probabilistic independence of object features to estimate conditional probability distributions [24]. The nearest-neighbour method is based on the not always correct assumption that if two objects are close to one another in terms of some distance measure then they belong, or are likely to belong to the same class. Experience indicates that nearest-neighbour techniques provide unsatisfactory performance and lose essential information in the process of reducing the classification problem to vector distance computation. Linear discriminant functions do not perform well in the general case when decision categories are not linearly separable [24, 27, 28]. The extension of linear discriminant functions, the multilayer back propagation neural nets, do not have this limitation and, in principle, can be trained to distinguish any two separable catogories [27]. They do, however, require tremendous computations power and the final result is strongly dependent on initial, intuitive parameter setting. Also, all these approaches suffer from dimensionality problems. They are not able to identify the essential subset of non-redundant factors (features), despite the fact that some dimensionality reduction algorithms have been developed. Whereas the performance of a rough-set-based system improves with the addition of extra relevant features, the performance of the statistical techniques can unexpectedly degrade when new features are added [24]. The use of the fuzzy set approach to approximate pattern classification has been advocated by Zadeh and his associates [22]. In this approach objects are represented in terms of features whose values are imprecise linguistic constants, such as, LOW, MEDIUM, etc., represented as fuzzy membership functions. The use of fuzzy membership functions allows one to express a degree of association of an object with an imprecisely defined linguistic notion. The classification result, the target class, is also assumed to be a fuzzy set rather than the precise set. The central problem in the fuzzy set approach is the derivation of a formula linking the membership functions of the unknown target class with given membership functions of object features. Given such a formula, for example, obtained from some "training" cases, one can predict the membership grades with respect to the target class in new cases. The formula can be perceived as a recognition rule corresponding to the set of logical rules forming target class descriptions in the rough set approach. The main difficulty with the fuzzy set method is the lack of objective techniques for defining feature or target class, membership functions.

3. Basic Notions and Properties

3.1. *Majority Inclusion Relation*

The heart of the extended rough set model is the generalization of the notion of the standard set inclusion relation. As we indicated in the Introduction, the extended notion should be able to allow for some degree of misclassification in the

largely correct classification. The standard definition of the set inclusion relation is too rigorous to represent any "almost" complete set inclusion.

Let X and Y be non-empty subsets of a finite universe U. We say that X is included in Y, or $Y \supseteq X$, if for all $e \in X$ implies $e \in Y$. Clearly, there is no room for even the slightest misclassification according to this definition. Therefore, before a more general definition is presented, it is convenient to introduce the measure $c(X, Y)$ of the relative degree of misclassification of the set X with respect to set Y defined as

$$c(X, Y) = 1 - \text{card}(X \cap Y)/\text{card}(X) \quad \text{if} \quad \text{card}(X) > 0 \text{ or}$$

$$c(X, Y) = 0 \quad \text{if} \quad \text{card}(X) = 0$$

where card denotes set cardinality.

That is, if we were to classify all elements of the set X into set Y then in $c(X, Y) * 100\%$ of the cases we would make a classification error. Consequently, the quantity $c(X, Y)$ will be referred to as the relative classification error. The actual number of misclassified elements is given by the product $c(X, Y) * \text{card}(X)$ which is referred to as an absolute classification error.

Based on the measure of relative misclassification one can define the inclusion relationship between X and Y without explicitly using a general quantifier:

$$Y \supseteq X \quad \text{if and only if} \quad c(X, Y) = 0.$$

The natural relaxation of this definition is to allow $c(X, Y)$ to assume values greater than 0. These values, however, cannot be too high if the relation is to represent a trend, or a requirement that a *specified majority* of objects in X be classified in Y.

The *majority* requirement implies that more than 50% of X elements should be in common with Y. The *specified majority* requirement imposes an additional restriction. The number of elements of X in common with Y should be above 50% and not below a certain limit, e.g., 85%. These requirements may be added to the extended definition of inclusion relation by specifying an explicit limitation on the admissible level of classification error. According to the specified majority requirement the admissible classification error β must be within the range $0 \leqslant \beta < 0.5$. Based on this assumption the majority inclusion relation is defined as

$$Y \overset{\beta}{\supseteq} X \quad \text{if and only if} \quad c(X, Y) \leqslant \beta.$$

The above definition covers the whole family of β-majority relations.

EXAMPLE 3.1. Let

$$X_1 = \{x_1, x_2, x_3, x_4\},$$

$$X_2 = \{x_1, x_2, x_5\},$$

$$X_3 = \{x_1, x_6, x_7\},$$

$$Y = \{x_1, x_2, x_3, x_8\}.$$

According to definition of the β-majority inclusion relation the following relationships are satisfied:

$$Y \overset{0.25}{\supseteq} X_1, \qquad Y \overset{0.33}{\supseteq} X_2,$$

but for every β it is not true that

$$Y \overset{\beta}{\supseteq} X_3.$$

It follows directly from the definition that the majority inclusion relation becomes a standard inclusion relation if $\beta = 0$. We will refer to standard inclusion relation as total inclusion.

It should be also noted that the majority inclusion relation does not have the transitivity property. Some useful properties of the majority inclusion relation are listed in Propositions 3.1 and 3.2.

PROPOSITION 3.1. *If* $A \cap B = \varnothing$ *and* $B \overset{\beta}{\supseteq} X$ *then it is not true that* $A \overset{\beta}{\supseteq} X$.

PROPOSITION 3.2. *If* $\beta_1 < \beta_2$ *then* $Y \overset{\beta_1}{\supseteq} X$ *implies* $Y \overset{\beta_2}{\supseteq} X$.

3.2. *Set Approximations in the VP-Model*

As in the original model of rough sets the approximation space is defined as a pair $A = (U, R)$ which consists of a non-empty, finite universe U and of the equivalence relation R on U. The equivalence relation R, referred to as an indiscernibility relation, corresponds to a partitioning of the universe U into a collection of equivalence classes or elementary sets $R^* = \{E_1, E_2, ..., E_n\}$. The central issue of the theory of rough sets is the specification of the discernibility limits of a set in U by means of elementary sets of R. As we illustrated in Section 2 the hypotheses derived from the standard set inclusion criterion may be too restrictive in some situations characterized by strong trends but the absence of total inclusion. By replacing the inclusion relation with a majority inclusion relation in the original definition of lower approximation of a set we obtain the following generalized notion of β-lower approximation or β-positive region, of the set $U \supseteq X$,

$$\underline{R}_\beta X = \bigcup \{E \in R^* : X \overset{\beta}{\supseteq} E\} \text{ or, equivalently,}$$

$$\underline{R}_\beta X = \bigcup \{E \in R^* : c(E, X) \leqslant \beta\}.$$

To follow the traditional notation of the theory of rough sets the β-lower approximation will also be called β-positive region of the set X and denoted alternatively as $\mathrm{POSR}_\beta(X)$.

The β-upper approximation of the set $U \supseteq X$ is defined as

$$\overline{R}_\beta X = \bigcup \{E \in R^* : c(E, X) < 1 - \beta\}$$

and, consequently, the β-boundary region of a set is given by

$$\mathrm{BNR}_\beta X = \bigcup \{E \in R^* : \beta < c(E, X) < 1 - \beta\}.$$

The β-negative region of X is defined as a complement of the β-upper approximation, i.e.,

$$\mathrm{NEGR}_\beta X = \bigcup \{E \in R^* : c(E, X) \geqslant 1 - \beta\}.$$

The lower approximation of the set X can be interpreted as the collection of all those elements of U which can be classified into X with the classification error not greater than β. Similarly, the β-negative region of X is the collection of all those elements of U which can be classified into the complement of X, $-X$ with the classification error not greater than β. The latter interpretation follows from the following simple Proposition 3.3.

PROPOSITION 3.3. *For every $X \subseteq U$ the following relationship is satisfied:*

$$\mathrm{POSR}_\beta(-X) = \mathrm{NEGR}_\beta X.$$

The β-boundary region of X consists of all those elements of U which cannot be classified either into X or into $-X$ with the classification error not greater than β. It should be interested to note here that the Law of Excluded Middle of propositional calculus does hold, in general, for imprecisely specified sets. The law states that, given proposition p, the disjunction $p \vee \neg p$ of p with its negation $\neg p$ is always true. For example, for any $x \in U$, and $X \subseteq U$ either $x \in X$ or $x \in -X$. Suppose now that set X is specified approximately in terms of its β-positive region $\mathrm{POSR}_\beta(X)$ and β-negative region $\mathrm{NEGR}_\beta(X)$. If the β-boundary region is empty, i.e., if $\mathrm{BNR}_\beta(X) = \varnothing$ then

$$\mathrm{POSR}_\beta(X) \cup \mathrm{NEGR}_\beta(X) = U.$$

Because β-positive and β-negative regions are disjoint then the proposition $x \in \mathrm{POSR}_\beta(X)$ or $x \in \mathrm{NEGR}_\beta(X)$ is true for any $x \in U$, indicating that in the case of an empty boundary region the Law of Excluded Middle will be satisfied for sets specified in terms of β-positive and β-negative regions. If, however , the β-boundary region is non-empty the proposition $x \in \mathrm{POSR}_\beta(X)$ or $x \in \mathrm{NEGR}_\beta(X)$ will not be true for all $x \in U$, as some x's would belong to $\mathrm{BNR}_\beta(X)$ rather than $\mathrm{POSR}_\beta(X)$ or $\mathrm{NEGR}_\beta(X)$, (the boundary region is disjoint with β-positive and negative regions). Consequently, the Law of Excluded Middle would not hold in the latter case.

Finally, the β-upper approximation $R_\beta X$ of X includes all those elements of U which cannot be classified into $-X$ with the error not greater than β.

By comparing the above definitions of set approximations with the definitions introduced in the original model of rough sets [1], one may notice that if $\beta = 0$ then the original rough set model becomes a special case of VP-model. This fact is summarized in the following proposition.

PROPOSITION 3.4. *Let X be an arbitrary subset of the universe U:*

1. $\underline{R}_0 X = \underline{R}X$, *where* $\underline{R}X$ *is a lower set approximation defined as* $\underline{R}X = \bigcup \{E \in R^*: X \supseteq E\}$

2. $\bar{R}_0 X = \bar{R}X$, *where* $\bar{R}X$ *is an upper set approximation defined as* $\bar{R}X = \bigcup \{E \in R^*: E \cap X \neq \emptyset\}$

3. $\mathrm{BNR}_0 X = \mathrm{BN}_R X$, *where* $\mathrm{BN}_R X$ *is the set X boundary region defined as* $\mathrm{BN}_R X = \bar{R}X - \underline{R}X$.

4. $\mathrm{NEGR}_0 X = \mathrm{NEG}_R X$ *where* $\mathrm{NEG}_R X$ *is the negative region defined as* $\mathrm{NEG}_R X = U - \bar{R}X$.

In addition to the properties listed in Proposition 3.4 for every $0 \leqslant \beta < 0.5$ the following relationships are also satisfied.

PROPOSITION 3.5.

$$\underline{R}_\beta X \supseteq \underline{R}X,$$

$$\bar{R}X \supseteq \bar{R}_\beta X,$$

$$\mathrm{BN}_R X \supseteq \mathrm{BNR}_\beta X,$$

$$\mathrm{NEGR}_\beta X \supseteq \mathrm{NEG}_R X.$$

Intuitively, with the decrease of the classification error β the size of the positive and negative regions of X will shrink, whereas the size of the boundary region will grow. With the reduction of β fewer elementary sets will satisfy the criterion for inclusion in β-positive or β-negative regions, thereby the size of the boundary will increase. Exactly the reverse process will occur with the increase of β. With the β approaching the limit 0.5, $\beta \to 0.5$, the set approximations will approach the following limits.

PROPOSITION 3.6.

$$\underline{R}_\beta X \to \underline{R}_{0.5} X = \bigcup \{E \in R^*: c(E, X) < 0.5\},$$

$$\bar{R}_\beta X \to \bar{R}_{0.5} X = \bigcup \{E \in R^*: c(E, X) \leqslant 0.5\},$$

$$\mathrm{BNR}_\beta \to \mathrm{BNR}_{0.5} X = \bigcup \{E \in R^*: c(E, X) = 0.5\},$$

$$\mathrm{NEGR}_\beta X \to \mathrm{NEGR}_{0.5} X = \bigcup \{E \in R^*: c(E, X) > 0.5\}.$$

The set $\mathrm{BNR}_{0.5} X$ is called an absolute boundary of X because it is included in every other boundary region of X. The following proposition summarizes the primary relationships between set X discernibility regions computed on 0.5 accuracy level and higher levels.

PROPOSITION 3.7.

$$\mathrm{BNR}_{0.5}X = \bigcap_{\beta} \mathrm{BNR}_{\beta}X,$$

$$\bar{R}_{0.5}X = \bigcap_{\beta} \bar{R}_{\beta}X,$$

$$\underline{R}_{0.5}X = \bigcup_{\beta} \underline{R}_{\beta}X,$$

$$\mathrm{NEG}_{0.5}X = \bigcup_{\beta} \mathrm{NEG}_{\beta}X.$$

The absolute boundary is, in a sense, very "narrow," consisting only of those elementary sets which have 50/50 aplit of elements among set X interior and its exterior. All other elementary sets are classified either into positive region $\underline{R}_{0.5}X$ or the negative region $\mathrm{NEGR}_{0.5}X$.

EXAMPLE 3.2. To illustrate the extended notions of set approximations consider the approximation space $A = (U, R)$ with $U = \{x_1, x_2, ..., x_{20}\}$ and the equivalence classes of the relation R given by:

$$E_1 = \{x_1, x_2, x_3, x_4, x_5\}$$

$$E_2 = \{x_6, x_7, x_8\}$$

$$E_3 = \{x_9, x_{10}. x_{11}, x_{12}\}$$

$$E_4 = \{x_{13}, x_{14}\}$$

$$E_5 = \{x_{15}, x_{16}, x_{17}, x_{18}\}$$

$$E_6 = \{x_{19}, x_{20}\}.$$

We will compute approximations of the set $X = \{x_4, x_5, x_8, x_{14}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}\}$ for two accuracy levels: $\beta_1 = 0$ and $\beta_2 = 0.25$.

If the assumed classification error $\beta = 0$, then, as it was noted in Proposition 3.9, the β-approximations of the set X are equal to the standard set approximations. Thus,

$$\underline{R}_0 X = E_6$$

$$\bar{R}_0 X = E_1 \cup E_2 \cup E_4 \cup E_5 \cup E_6$$

$$\mathrm{BNR}_0 X = E_1 \cup E_2 \cup E_4 \cup E_5$$

$$\mathrm{NEGR}_0 X = E_3.$$

As we relax the admissible classification error to become $\beta_2 = 0.25$ then the negative and positive regions start to grow at the expense of the boundary region, leading to the following result:

$$\underline{R}_{0.25}X = E_5 \cup E_6$$

$$\bar{R}_{0.25}X = E_1 \cup E_2 \cup E_4 \cup E_5 \cup E_6$$

$$\mathrm{BNR}_{0.25}X = E_1 \cup E_2 \cup E_4$$

$$\mathrm{NEGR}_{0.25}X = E_3.$$

3.3. *Measure of Approximation*

To express the degree with which a set X can be approximately characterized by means of elementary sets of the approximation space $A = (U, R)$, we will generalize the accuracy measure introduced in [1]. The β-accuracy for $0 \leqslant \beta < 0.5$ is defined as

$$\alpha(R, \beta, X) = \mathrm{card}(\underline{R}_\beta X)/\mathrm{card}(\bar{R}_\beta X).$$

The β-accuracy represents the imprecision of the approximate characterization of the set X relative to assumed classification error β. It is interesting to note that with the increase of β the cardinality of the β-upper approximation will tend downward and the size of the β-lower approximation will tend upward which leads to the conclusion that is consistent with intuition that the relative accuracy may increase at the expense of a higher classification error.

3.4. *Relative Discernability of Sets*

The notion of discernability of set boundaries is relative. If a large classification error is allowed then the set X can be highly discernable within assumed classification error limits. When smaller values of the classification tolerance are assumed it may become more difficult to discern positive and negative regions of the set to meet the narrow tolerance limits.

The set X is said to be β-discernable if its β-boundary region is empty or, equivalently, if

$$\underline{R}_\beta X = \bar{R}_\beta X.$$

For the β-discernible sets the relative accuracy $\alpha(R, \beta, X)$ is equal to unity. The discernibility status of a set can change depending on the value of β. In general, the following properties are true.

PROPOSITION 3.8. *If X is discernible on the classification error level $0 \leqslant \beta < 0.5$ then X is also discernible at any level $\beta_1 > \beta$.*

PROPOSITION 3.9. *If $\bar{R}_{0.5}X \neq \underline{R}_{0.5}X$ then the set X is not discernible on every classification error level $0 \leqslant \beta < 0.5$.*

The latter proposition emphasizes the fact that a set with a nonempty absolute boundary can never be discerned. In general, one can easily demonstrate the following.

PROPOSITION 3.10. *If X is not discernible on the classification error level $0 \leqslant \beta < 0.5$ then X is also not discernible at any level $\beta_2 < \beta$.*

Any set X which is not discernible for every β will be called absolutely indiscernible or absolutely rough. The set X is absolutely rough if and only if $\mathrm{BNR}_{0.5} X \neq \varnothing$. Any set which is not absolutely rough will be referred to as relatively rough or weakly discernible. For each relatively rough set X there exists such a classification error level β that set X is discernible on this level.

Let $\mathrm{NDIS}(R, X) = \{0 \leqslant \beta < 0.5 \colon \mathrm{BNR}_\beta(X) \neq \varnothing\}$. $\mathrm{NDIS}(R, X)$ is a range of all those β values for which X is indiscernible. The least value of classification error β which makes X discernable will be referred to as discernibility threshold. The value of the threshold is equal to the least upper bound $\zeta(R, X)$ of $\mathrm{NDIS}(X)$, i.e.,

$$\zeta(R, X) = \sup \mathrm{NDIS}(R, X).$$

Proposition 3.11 provides a simple property which can be used to find the discernibility threshold of a weakly discernible set X.

PROPOSITION 3.11. $\zeta(R, X) = \max(m_1, m_2)$, *where*

$$m_1 = 1 - \min\{c(E, X) \colon E \in R^* \ \& \ 0.5 < c(E, X)\},$$
$$m_2 = \max\{c(E, X) \colon E \in R^* \ \& \ c(E, X) < 0.5\}.$$

EXAMPLE 3.3. To illustrate the latter point let us assume that, given the following elementary classes,

$$E_1 = \{x_1, x_2, x_3, x_4, x_5\},$$
$$E_2 = \{x_6, x_7, x_8\},$$
$$E_3 = \{x_9, x_{10}, x_{11}, x_{12}\},$$
$$E_4 = \{x_{13}, x_{14}, x_{15}, x_{16}\},$$
$$E_5 = \{x_{17}, x_{18}\}, \text{ and the set}$$
$$X = \{x_4, x_5, x_8, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}\},$$

the classification errors of the set X computed for all classes are:

$$c(E_1, X) = 0.6,$$
$$c(E_2, X) = 0.66,$$
$$c(E_3, X) = 1.0,$$
$$c(E_4, X) = 0.25,$$
$$c(E_5, X) = 0.0.$$

It can be easily verified, based on the Proposition 3.11, that the least value for which X is still discernible is $\zeta(R, X) = 0.4$.

The discernibility threshold of the set X equals a minimal classification error β which must be allowed in order to make this set β-discernible.

4. Properties of β-Approximations

In this section we will demonstrate some fundamental properties of β-approximations. With a few exceptions these properties are identical to properties of approximations given in [16].

Proposition 4.1. *For every* $0 \leqslant \beta < 0.5$ *the following relationships are true*:

(1a) $X \overset{\beta}{\supseteq} \underline{R}_\beta X$

(1b) $\bar{R}_\beta X \supseteq \underline{R}_\beta X$

(2) $\underline{R}_\beta \varnothing = \bar{R}_\beta \varnothing = \varnothing$; $\underline{R}_\beta U = \bar{R}_\beta U = U$

(3) $\bar{R}_\beta(X \cup Y) \supseteq \bar{R}_\beta X \cup \bar{R}_\beta Y$

(4) $\underline{R}_\beta X \cap \underline{R}_\beta Y \supseteq \underline{R}_\beta(X \cap Y)$

(5) $\underline{R}_\beta(X \cup Y) \supseteq \underline{R}_\beta X \cup \underline{R}_\beta Y$

(6) $\bar{R}_\beta X \cap \bar{R}_\beta Y \supseteq \bar{R}_\beta(X \cap Y)$

(7) $\underline{R}_\beta(-X) = -\bar{R}_\beta(X)$

(8) $\bar{R}_\beta(-X) = -\underline{R}_\beta X$.

Proof. (1a) To demonstrate that $X \overset{\beta}{\supseteq} \underline{R}_\beta X$ it suffices to show that for any two elementary sets E_1, E_2 if $c(E_1, X) \leqslant \beta$ and $c(E_2, X) \leqslant \beta$ then

$$c(E_1 \cup E_2, X) \leqslant \beta.$$

Directly from the definition of the measure $c(X, Y)$ we obtain

(i) $c(E_1, X) = \text{card}(E_1 \cap (-X))/(\text{card}(E_1 \cap (-X)) + \text{card}(E_1 \cap X)) \leqslant \beta$

(ii) $c(E_2, X) = \text{card}(E_2 \cap (-X))/(\text{card}(E_2 \cap (-X)) + \text{card}(E_2 \cap X)) \leqslant \beta$.

From (i) and (ii) it follows that

$$\begin{aligned} &\text{card}(E_1 \cap (-X)) + \text{card}(E_2 \cap (-X)) \\ &\quad \leqslant \beta(\text{card}(E_1 \cap (-X)) + \text{card}(E_1 \cap X) \\ &\qquad + \text{card}(E_2 \cap (-X)) + \text{card}(E_2 \cap X)) \end{aligned}$$

which means that

$$c(E_1 \cup E_2, X) = \text{card}((E_1 \cup E_2) \cap (-X))/\text{card}(E_1 \cup E_2) \leqslant \beta.$$

(1b) The inclusion $\bar{R}_\beta X \supseteq \underline{R}_\beta X$ follows directly from definitions of lower and upper approximations.

(2) Since $c(E, \varnothing) = 1$ for any elementary set E, it follows that $\underline{R}_\beta \varnothing = \varnothing$ and $\bar{R}_\beta \varnothing = \varnothing$. Similarly, since $c(E, U) = 0$ it follows that $\underline{R}_\beta(U) = U$ and $\bar{R}_\beta(U) = U$.

(3) This property follows from the simple fact that if $U \supseteq X$, Y then $c(E, X \cup Y) \leqslant c(E, X)$ and $c(E, X \cup Y) \leqslant c(E, Y)$.

(4) This inclusion is a direct consequence of the relationship $c(E, X \cap Y) \geqslant c(E, X)$ and $c(E, X \cap Y) \geqslant c(E, Y)$.

(5) This property follows again from the fact given in (3).

(6) Same as (4).

(7) Can be simply derived from the property $c(E, -X) = 1 - c(E, X)$.

(8) Also directly follows from the formula given in (7) and definitions of lower and upper bounds.

One essential difference in comparison to the group of properties listed in [16] for the standard rough set model is the absence of the inclusion $\bar{R}_\beta X \overset{\beta}{\supseteq} X$ in the generalized case. This is due to the fact that for any positive β, the β-negative region $\mathrm{NEG}_\beta X$ is not, in general, disjoint from X.

5. Analysis of Information Systems

The primary motivation behind the development of the extended approach to rough sets is the need for a more flexible tool for analysis of information systems. The logical notion of an information system was introduced in [8] to formalize basic components of such a knowledge representation technique in which information is expressed by attributes and their values. As it will be demonstrated below it is possible to use attribute values to impose a structure of an approximation space on such an information system. This, in turn, leads to the important definition of attribute dependency which is based on the idea of set approximations. The attribute dependency measure expresses the degree of functional relationship among two groups of attributes, for example, symptoms versus diseases in a medical application. The computation of attribute dependency combined with attribute reduction and attribute weight computation provides an analytical tool for a logical analysis of properties of data.

The major difficulty associated with the analytical methodology based on rough sets is the high sensitivity of computational results to small misclassification errors. This is caused by the fact that the definition of attribute dependency [16] is based on the standard notion of set inclusion which excludes any misclassification errors. The generalization of this definition to allow for some degree of misclassification is derived from the notion of partial inclusion relation.

5.1. *Knowledge Tables*

According to the definition given in [8] a knowledge representation system, or an information system, is a quadruple

$S=(U, A, V, f)$ such that

U is the universe of objects,

A is the collection of object attributes,

V is the union of attribute domains, i.e., $V=\bigcup V_a$ for $a \in A$, where V_a denotes the domain of the attribute a and

f is an information function which associates a unique value of each attribute with every object belonging to U.

An information system can be conveniently represented in the form of a knowledge table in which rows correspond to objects represented by attribute values. An example of a knowledge table is shown in Table I.

In this section we will focus on analysis of the relationship between two groups of attributes $A \supseteq P$, Q referred to respectively as condition and decision attributes.

EXAMPLE 5.1. In the Table I a collection of cars is described in terms of attributes such as overall length (size), number of cylinders (cylinder), presence of a turbocharger (turbo), type of fuel system (fuelsys), engine displacement (displace), compression ratio (compress), type of transmission (transmis), and gas mileage. The information was collected from car test results published by *Popular Science*. To analyze such an information system with respect to the relationship between car gas mileage and other parameters, the mileage and other parameters, the mileage attribute has been declared as a decision attribute Q and all remaining attributes form a set of conditions P.

5.2. *Approximate Dependency of Attributes*

The approximate dependency measure of attributes is defined, based on the idea of lower approximation of a set. Let IND(P) and IND(Q) be two indiscernibility relations imposed on U by sets of attributes P and Q according to the following definition:

> Two objects $x, y \in U$ are equivalent with respect to P, i.e., $(x, y) \in \text{IND}(P)$ if and only if $f(x, a)=f(y, a)$ for all $a \in P$.

In other words, two objects are said to be equivalent if they have identical values of attributes belonging to P. The set of equivalence classes of the relation IND(P) will be referred to as condition classes and denoted as P^*. Similarly, the collection of equivalence classes of the relation IND(Q) will be called decision classes and denoted as Q^*.

TABLE I

U	Size	Cylinder	Turbo	Fuelsys	Displace	Compress	Power	Transmis	Weight	Mileage
1	Compact	6	Yes	EFI	Medium	High	High	Auto	Medium	Medium
2	Compact	6	No	EFI	Medium	Medium	High	Manual	Medium	Medium
3	Compact	6	No	EFI	Medium	High	High	Manual	Medium	Medium
4	Compact	4	Yes	EFI	Medium	High	High	Manual	Light	High
5	Compact	6	No	EFI	Medium	Medium	Medium	Manual	Medium	Medium
6	Compact	6	No	2-bbl	Medium	Medium	Medium	Auto	Heavy	Low
7	Compact	6	No	EFI	Medium	Medium	High	Manual	Heavy	Low
8	Subcompact	4	No	2-bbl	Small	High	Low	Manual	Light	High
9	Compact	4	No	2-bbl	Small	High	Low	Manual	Medium	Medium
10	Compact	4	No	2-bbl	Small	High	Medium	Auto	Medium	Medium
11	Subcompact	4	No	EFI	Small	High	Low	Manual	Light	High
12	Subcompact	4	No	EFI	Medium	Medium	Medium	Manual	Medium	High
13	Compact	4	No	2-bbl	Medium	Medium	Medium	Manual	Medium	Medium
14	Subcompact	4	Yes	EFI	Small	High	High	Manual	Medium	High
15	Subcompact	4	No	2-bbl	Small	Medium	Low	Manual	Medium	High
16	Compact	4	Yes	EFI	Medium	Medium	High	Manual	Medium	Medium
17	Compact	6	No	EFI	Medium	Medium	High	Auto	Medium	Medium
18	Compact	4	No	EFI	Medium	Medium	High	Auto	Medium	Medium
19	Subcompact	4	No	EFI	Small	High	Medium	Manual	Medium	High
20	Compact	4	No	EFI	Small	High	Medium	Manual	Medium	High
21	Compact	4	No	2-bbl	Small	High	Medium	Manual	Medium	Medium

The set of attributes Q is said to β-depend in degree $\gamma(P, Q, \beta)$ on the set of attributes P if

$$\gamma(P, Q, \beta) = \text{card POS}(P, Q, \beta)/\text{card } U,$$

where $\text{POS}(P, Q, \beta)$ is a β-positive region of the partition Q^* defined as $\text{POS}(P, Q, \beta) = \bigcup_{Y \in Q^*} \underline{\text{IND}}(P)_\beta Y$.

The β-dependency level is a measure of the relative size of the union of β-positive regions of all equivalence classes of the relation $\text{IND}(Q)$ computed in the approximation space induced by $\text{IND}(P)$. The β-dependency, when computed from experimental data, can be interpreted as a proportion of those cases in U which can be discriminated between different classes of the relation $\text{IND}(Q)$ with the classification error less than $0 \leqslant \beta < 0.5$.

The approximate dependency is clearly a generalization of the idea of rough dependency [16] as it becomes rough dependency for $\beta = 0$. Informally speaking, the rough dependency measure is an evaluation of the overall ability to perform exact, i.e., error free classification of objects, whereas the approximate dependency measures the ability to do the classification with an error falling into the preset tolerance limit β. The approximate dependency, as opposed to the rough dependency, cannot be interpreted as a functional or partial functional dependency of attributes. The properties of approximate dependencies are much weaker than properties of functional dependencies; for instance, the transitivity property does not hold.

EXAMPLE 5.2. To illustrate the idea of approximate dependency we will compute the dependency level between attributes $P = \{a, b, c\}$ and $Q = \{d\}$ for different values of the tolerance limit β based on Table II.

There are four classes of the partition P^*:

$X_1 = \{1, 2, 19, 20, 21\}$, $X_2 = \{3\}$, $X_3 = \{4\text{–}13\}$,
$X_4 = \{14\text{–}18\}$ and there are three classes $Y_1 = \{1\text{–}12\}$,
$Y_2 = \{13\text{–}17\}$, $Y_3 = \{18\text{–}21\}$ of the partition Q^*.

The classes of the partition Q^* are represented by values of the attribute d in the Table II.

Let us assume first the error tolerance level $\beta = 0$. To compute the positive region of Q^* we need to find all those condition classes which are completely included in some decision classes. The only such class is X_2 and therefore to 0-dependency, or rough dependency, is

$$\gamma(P, Q, 0) = \text{card}(X_2)/\text{card}(U) = 1/21 = 0.047.$$

For the increased tolerance $\beta = 0.1$ to compute the 0.1-positive region of Q^* one

TABLE II

U	a	b	c	d
1	1	0	0	1
2	1	0	0	1
3	1	1	1	1
4	0	1	1	1
5	0	1	1	1
6	0	1	1	1
7	0	1	1	1
8	0	1	1	1
9	0	1	1	1
10	0	1	1	1
11	0	1	1	1
12	0	1	1	1
13	0	1	1	2
14	1	1	0	2
15	1	1	0	2
16	1	1	0	2
17	1	1	0	2
18	1	1	0	3
19	1	0	0	3
20	1	0	0	3
21	1	0	0	3

has to identify all classes of P^* which are included in some classes of Q^* with an error not greater than 0.1. There are two classes which satisfy this condition:

$$Y_1 \overset{0.1}{\supseteq} X_2 \qquad \text{and} \qquad Y_1 \overset{0.1}{\supseteq} X_3.$$

From these two classes we can compute the 0.1-dependency measure between attributes P and Q as

$$\gamma(P, Q, 0.1) = \text{card}(X_2 \cup X_3)/\text{card } U = 11/21 = 0.52.$$

Similarly, if β is set to 0.2 then

$$Y_1 \overset{0.2}{\supseteq} X_2, \qquad Y_1 \overset{0.2}{\supseteq} X_3, \qquad Y_2 \overset{0.2}{\supseteq} X_4$$

which result in the dependency level

$$\gamma(P, Q, 0.2) = \text{card}(X_2 \cup X_3 \cup X_4)/\text{card } U = 16/21 = 0.76.$$

5.3. *Approximate Reduct*

One of the most important notions of the rough sets model is the notion of attribute reduct or minimal set of attributes. According to the definition given in [16], a reduct is the minimal subset of condition attributes P preserving the

dependency with decision attributes Q. By adapting the extended definition of approximate dependency one can generalize the idea of a reduct by introducing the approximate reduct as follows.

A β-reduct, or approximate reduct, of the set of condition attributes P with respect to a set of decision attributes Q is a subset $\mathrm{RED}(P, Q, \beta)$ of P which satisfies the following two criteria:

(1) $\gamma(P, Q, \beta) = \gamma(\mathrm{RED}(P, Q, \beta), Q, \beta)$

(2) no attribute can be eliminated from $\mathrm{RED}(P, Q, \beta)$ without affecting the requirement (1).

The idea of reduct turned out to be the most useful in those applications where it was necessary to find the most important collection of attributes responsible for a cause-and-effect relationship. It is also useful for eliminating noise attributes from the table of observations. The following example illustrates the computation of reducts.

EXAMPLE 5.3. There are over 100 different minimal sets of attributes which can be computed from the information system presented in the Table I with condition attributes P and the decision attribute Q defined as in Example 5.1 (the tolerance level is $\beta = 0$). Some of them are shown below:

1. cylinder, fuelsys, compress, power, weight
2. size, fuelsys, compress, power, weight
3. size, fuelsys, displace, weight
4. size, cylinder, fuelsys, power, weight
5. cylinder, turbo, fuelsys, displace, compress, weight
6. size, cylinder, fuelsys, compress, weight
7. size, cylinder, turbo, fuelsys, transmis, weight
8. size, displace, weight, price.

Each of the above minimal sets of attributes can be used to represent information about cars *instead* of all the condition attributes.

5.4. *Selection of the Best Minimal Set of Attributes*

Every minimal set of attributes may be perceived as an alternative group of attributes which could be used instead of all available attributes in the decision making based on cases. The main difficulty is how to select an optimal reduct. This selection depends on the optimality criterion associated with attributes. If it is possible to assign a cost function to attributes then the selection can be based naturally on the combined minimum cost criterion. For example, in the medical domain, some diagnostic procedures are much more expensive than the others. By selecting the least expensive series of tests represented by the minimum cost reduct,

considerable savings can be accomplished without decreasing the quality of diagnosis. In the absence of attribute cost function the only source of information to select the reduct is the contents of the table. Two approaches are possible in this case. In the first one, the reduct with the minimal number of attributes is selected. In the second approach, the reduct which has the least number of combinations of values of its attributes is selected. The latter criterion favors reduct which represents the strongest pattern or data regularity.

For instance, in the collection of reducts presented in Example 5.3, the best reduct is

8. size, displace, weight, price,

which has only 10 combinations of attribute values, while other reducts have approximately 20 combinations.

6. Closing Remarks

The variable precision rought set model presented in this paper is a direct generalization of the original rough sets model published in [1]. Since its introduction over 300 research papers dealing with theoretical as well as practical aspects of the rough sets model have been published across the world. Several software systems aimed at machine learning and data analysis applications of rough sets were developed [3, 7, 12, 15, 17]. For example, all computational results presented in this paper were produced by using a commercial software package for knowledge acquisition called DataQuest. Despite this significant progress the area is still not mature in either the theoretical or the practical area. The generalization discussed in this paper illustrates that there is much room for expansion of the theory. To demonstrate the practical utility of the approach, further research is needed. In particular, there is a need for objective comparative studied classification results produced by rough set-based algorithms versus other methods such as statistical, neural nets, etc. Practical implementations of the rough set model and the investigation of its utility in different areas of engineering and science are very costly and time consuming. Therefore, one cannot expect quick growth on the applications side of the rough sets methodology despite continuous efforts in this direction. The broad availability of general purpose commercial software packages for data analysis and rule asquisition with rough sets will likely accelerate the practical utilization of the model. Also, a comprehensive monography describing the fundamentals of the original rough set model [29] and a book describing selected practical applications of rough sets [30] published recently will contribute to better understanding of the model and to the development of new applications.

ACKNOWLEDGMENTS

Thanks are due to Dr. Zdzislaw Pawlak and Dr. Jack Katzberg for comments and suggestions leading to the improved version of the paper. The reported research was supported in part by an operating grant from Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. Z. PAWLAK, Rough Sets, *Internat. J. Comput. Inform. Sci.* **11**, No. 5 (1982), 341–356.
2. H. RASIOWA, Rough concepts and multiple valued logic, *in* "Proceedings, 16th International Symposium on Multiple Valued Logic," pp. 228–288, Computer Society Press, New York, 1986.
3. G. EPSTEIN *et al.*, Approximation reasoning and Scott's information systems, *in* "Proceedings, 2nd International Symposium on Methodologies for Intelligent Systems," North Holland, Amsterdam, 1987.
4. A. MROZEK, Rough sets and some aspects of expert system realization, *in* "Proceedings, 7th International Workshop on Expert Systems, Avignon, France, 1987," pp. 587–611.
5. A. MROZEK, Rough sets and dependency analysis among attributes in computer implementation of experts inference models, *Internat. J. Man-Mach. Stud.* **30**, No. 4 (1989), 448–457.
6. J. GRZYMALA-BUSSE, Knowledge acquisition under uncertainty—A rough set approach, *J. Intell. Rob. Systems* **1** (1988), 3–16.
7. J. GRZYMALA-BUSSE AND D. J. SIKORA, "LERS1—A System for Learning from Examples Based on Rough Sets," Tech. Rep., TR-88-5, University of Kansas, 1988.
8. Z. PAWLAK AND E. ORLOWSKA, Expressive power of knowledge representation, *Internat. J. Man-Mach. Stud.* **20** (1984), 485–500.
9. A. WASILEWSKA, Syntactic decision procedures in information systems, *Internat. J. Man-Mach. Stud.* **30** (1989), 273–285.
10. R. YASDI AND W. ZIARKO, An expert system for conceptual scheme design: A machine learning approach, *Internat. J. Man-Mach. Stud.* **29** (1988), 351–376.
11. S. K. M. WONG, W. ZIARKO, AND Y. LI, Comparison of rough set and statistical methods in inductive learning, *Internat. J. Man-Mach. Stud.* **24** (1986), 53–72.
12. W. ZIARKO, A technique for discovering and analysis of cause-effect relationships in empirical data, *in* "International Joint Conference on Artificial Intelligence, Proceedings, Workshop on Knowledge Discovery in Databases, Detroit, 1989."
13. W. ZIARKO AND S. K. M. WONG, INFER – An adaptive decision support system, *in* "Proceedings of the 6th International Workshop on Expert Systems and Their Applications, Avignon, France, 1986," pp. 713–726.
14. S. K. M. WONG AND W. ZIARKO, A machine learning approach to information retrieval, *in* "Proceedings, 9th International ACM-SIGIR Conference on Research and Development in Information Retrieval, Pisa, Italy, 1986," pp. 228–233.
15. W. ZIARKO, Data analysis and case-based expert system development tool ROUGH, *in* "Proceedings, Workshop on Case-Based Reasoning, Pensacola Beach, Florida," Morgan Kaufmann, Los Altos, CA, 1989.
16. Z. PAWLAK, S. K. M. WONG, AND W. ZIARKO, Rough sets, probabilistic versus deterministic approach, *Internat. J. Man-Mach. Stud.* **29** (1988), 81–95.
17. H. KRASOWSKI, "Rough Set-Based Method for Evaluation of Pilot Performance," Ph.D. thesis, Technical University of Rzeszow, 1988.
18. A. KOWALCZYK AND J. SZYMANSKI, "Rough Simplification of Decision Tables," Telecom Tech. Branch Paper 88/17, Australian Telecommunication Commission, 1988.
19. Z. PAWLAK, K. SLOWINSKI, AND R. SLOWINSKI, Rough classification of patients after highly selective vogotomy for duodenal ulcer, *Internat. J. Man-Mach. Stud.* **24** (1986), 413–433.

20. P. HAJDO, T. ARCISZEWSKI, W. ZIARKO, AND H. AKTAN, Inductive shallow approach for generation of engineering models, *in* "Proceedings, 9th European Meeting on Cybernetics, and Systems Research, Vienna, Austria, 1988," pp. 933–940.
21. W. ZIARKO AND J. KATZBERG, Control algorithm acquisition, analysis and reduction: A machine learning approach, *in* "Knowledge-Based System Diagnosis, Supervision and Control," pp. 167–178, Plenum, New York, 1989.
22. L. A. ZADEH, Fuzzy sets and their applications to pattern classification and clustering analysis, *in* "Classification and Clustering (J. Van Ryzin, Ed.), Academic Press, New York/London, 1977.
23. I. H. LAVALLE, "An Introduction to Probability, Decision and Inference," Holt, Rinehart, & Winston, New York, 1970.
24. R. O. DUDA AND P. E. HART, "Pattern Classification and Scene Analysis," Wiley, New York, 1973.
25. W. M. MEISEL, "Computer-Oriented Approaches to Pattern Recognition," Academic Press, New York/London, 1972.
26. C. CHEN, "Statistical Pattern Recognition," Hayden, New York, 1973.
27. S. M. WEISS AND C. A. KULIKOWSKI, "Computer Systems That Learns," Morgan Kaufmann, San Mateo, CA, 1991.
28. M. J. NILSSON, "The Mathematical Foundations of Learning Machines," Morgan Kaufmann, San Mateo, CA, 1990.
29. Z. PAWLAK, "Rough Sets: Theoretical Aspects of Reasoning About Data," Kluwer Academic, Dordrecht, The Netherlands, 1991.
30. R. SLOWINSKI, (ed.), "Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory," Kluwer Academic, Boston, 1992.