In Search of Effective Granulization with DTRS for Ternary Classification

Bing Zhou, University of Regina, Canada

Yiyu Yao, University of Regina, Canada

ABSTRACT

Decision-Theoretic Rough Set (DTRS) model provides a three-way decision approach to classification problems, which allows a classifier to make a deferment decision on suspicious examples, rather than being forced to make an immediate determination. The deferred cases must be reexamined by collecting further information. Although the formulation of DTRS is intuitively appealing, a fundamental question that remains is how to determine the class of the deferment examples. In this paper, the authors introduce an adaptive learning method that automatically deals with the deferred examples by searching for effective granulization. A decision tree is constructed for classification. At each level, the authors sequentially choose the attributes that provide the most effective granulization. A subtree is added recursively if the conditional probability lies in between of the two given thresholds. A branch reaches its leaf node when the conditional probability is above or equal to the first threshold, or is below or equal to the second threshold, or the granule meets certain conditions. This learning process is illustrated by an example.

Keywords: *Decision-Theoretic Rough Sets, Granulization, Subtree, Ternary Classification, Three-Way Decision Approach*

INTRODUCTION

The Decision-Theoretic Rough Set (DTRS) model, proposed by Yao et al. (Yao, Wang, & Lingras, 1990; Yao & Wang, 1992; Yao, 2010) in the early 1990s, is a meaningful and useful generalization of the probabilistic rough set model (Pawlak, 1991). In probabilistic rough set models, three probabilistic regions are defined by considering the degree of overlap between an equivalence class and a set to be approximated. A conditional probability is used to state the degree of overlap and a pair of thresholds is used to define the three regions. An equivalence class is in the probabilistic positive region if its relative overlap with the set is above or equal to a threshold, is in the negative region if its relative overlap is below or equal to another threshold, and is in the boundary region if the relative overlap is between the two parameters. DTRS offers a solid

DOI: 10.4018/jcini.2011070103

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

foundation for probabilistic rough sets by systematically calculating the pair of thresholds based on the well-established Bayesian decision theory. Many real world problems can be solved with DTRS. For instance, DTRS provides a three-way decision approach to classification problems by allowing the possibility of indecision to suspicious examples, those examples in the boundary region must be re-examined by collecting additional information. A fundamental question that remains in DTRS is how to determine the classification of these deferred examples.

Cognitive science and cognitive informatics (Wang, 2007; Wang et al., 2009, 2011) study the human intelligence and its computational process. As an effective way of thinking, we typically focus on a particular level of abstraction and ignore irrelevant details. This not only enables us to identify differences between objects in the real world, but also helps us to view different objects as being the same, if low-level detail is ignored. Granular computing (GrC) (Bargiela & Pedrycz, 2002; Liang & Qian, 2008; Qian, Liang, & Dang, 2009; Yao, 2004b, 2007b, 2009) can be seen as a formal way of modeling this human thinking process. GrC is an area of study that explores different levels of granularity in human-centered perception, problem solving, and information processing, as well as their implications and applications in the design and implementation of knowledge intensive intelligent systems. Rough set theory is one of the concrete models of GrC for knowledge representation and data analysis.

In this paper, an adaptive learning method is introduced that classifies the deferred examples by adaptively searching for effective granulization. A decision tree is constructed for classification. At each level, we sequentially choose the attributes that provide the most suitable granulization. A subtree is added if the conditional probability lies in between of the two thresholds. A branch reaches its leaf node when the conditional probability is above or equal to the first threshold, or is below or equal to the second threshold.

The rest of the paper is organized as follows. We briefly review the basic ideas of DTRS. We introduce the interpretations of concepts based on GrC. A new adaptive learning algorithm is introduced for ternary classification. An illustrative example is given. We conclude the paper and explain the future work.

BRIEF INTRODUCTION TO DECISION-THEORETIC ROUGH SET MODEL

Bayesian decision theory is a fundamental statistical approach that makes decisions under uncertainty based on probabilities and costs associated with decisions. Following the discussions given in the book by Duda and Hart (1973), the decision theoretic rough set model is a straightforward application of the Bayesian decision theory.

With respect to the set C to be approximated, we have a set of two states $\Omega = \{C, C^C\}$ indicating that an object is in C or not in C, respectively. We use the same symbol to denote both a set C and the corresponding state. With respect to the three regions in the rough set theory, the set of actions is given by $\mathbf{A} = \{a_P, a_B, a_N\}$, where a_P, a_B and a_N represent the three actions in classifying an object x, namely, deciding $x \in \text{POS}(C)$, deciding $x \in \text{BND}(C)$, and deciding $x \in \text{NEG}(C)$, respectively. The loss function is given by the 3x2 matrix.

In the matrix, λ_{pp}, λ_{BP} and λ_{NP} denote the losses incurred for taking actions a_P, a_B and a_N respectively, when an object belongs to C, and λ_{PN}, λ_{BN}, and λ_{NN} denote the losses incurred for taking these actions when the object does not belong to C.

We use $\Pr(C|[x])$ to represent the conditional probability of an object belonging to C given that the object is described by its equivalence class $[x]$. The expected losses associated with taking different actions for objects in $[x]$ can be expressed as:

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Matrix 1.

	C(P)	$C^C(N)$
a_P	$\lambda_{PP} = \lambda(a_P \mid C)$	$\lambda_{PN} = \lambda(a_P \mid C^C)$
a_B	$\lambda_{BP} = \lambda(a_B \mid C)$	$\lambda_{BN} = \lambda(a_B \mid C^C)$
a_N	$\lambda_{NP} = \lambda(a_N \mid C)$	$\lambda_{NN} = \lambda(a_N \mid C^C)$

$$R(a_P \mid [x]) = \lambda_{PP} \Pr(C \mid [x]) + \lambda_{PN} \Pr(C^C \mid [x]),$$
$$R(a_B \mid [x]) = \lambda_{BP} \Pr(C \mid [x]) + \lambda_{BN} \Pr(C^C \mid [x]),$$
$$R(a_N \mid [x]) = \lambda_{NP} \Pr(C \mid [x]) + \lambda_{NN} \Pr(C^C \mid [x]).$$

The Bayesian decision procedure suggests the following minimum-risk decision rules:

(P) If: $R(a_P \mid [x]) \le R(a_B \mid [x]) \,\&\, R(a_P \mid [x]) \le R(a_N \mid [x])$, decide $x \in POS(C)$;
(B) If : $R(a_B \mid [x]) \le R(a_P \mid [x]) \,\&\, R(a_B \mid [x]) \le R(a_N \mid [x])$, decide $x \in BND(C)$;
(N) If : $R(a_N \mid [x]) \le R(a_P \mid [x]) \,\&\, R(a_N \mid [x]) \le R(a_B \mid [x])$, decide $x \in NEG(C)$.

Tie-breaking criteria should be added so that each object is put into only one region.

Since Pr(C|[x]) + Pr(C^C |[x]) = 1, we can simplify the rules based only on the probabilities Pr(C|[x]) and the loss function λ. Consider a special kind of loss functions with:

(c0). $\lambda_{PP} \le \lambda_{BP} < \lambda_{NP,}\ \lambda_{NN} \le \lambda_{BN} < \lambda_{PN}$.

That is, the loss of classifying an object x belonging to C into the positive region POS(C) is less than or equal to the loss of classifying x into the boundary region BND(C), and both of these losses are strictly less than the loss of classifying x into the negative region NEG(C). The reverse order of losses is used for classifying an object not in C. Under condition (c0), the decision rules (P)-(N) can be re-expressed as:

(P) If $\Pr(C \mid [x]) \ge \alpha \,\&\, \Pr(C \mid [x]) \ge \gamma$, decide $x \in POS(C)$;
(B) If $\Pr(C \mid [x]) \le \alpha \,\&\, \Pr(C \mid [x]) \ge \beta$, decide $x \in BND(C)$;
(N) If $\Pr(C \mid [x]) \le \beta \,\&\, \Pr(C \mid [x]) \le \gamma$, decide $x \in NEG(C)$;

where the threshold values α, β and γ are given by:

$$\alpha = \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})},$$
$$\beta = \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})},$$

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

$$\gamma = \frac{(\lambda_{PN} - \lambda_{NN})}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})} .$$

In other words, from a loss function one can systematically determine the required threshold values. When $(\lambda_{PN} - \lambda_{BN})(\lambda_{NP} - \lambda_{BP}) > (\lambda_{BP} - \lambda_{PP})(\lambda_{BN} - \lambda_{NN})$, we have $\alpha > \beta$, and thus $\alpha > \gamma > \beta$, after tie-breaking, we obtain:

P1. If $\Pr(C \mid [x]) \geq \alpha,$, decide $x \in POS(C)$;
B1. If $\beta < \Pr(C \mid [x]) < \alpha,$, decide $x \in BND(C)$;
N1. If $\Pr(C \mid [x]) \leq \beta,$, decide $x \in NEG(C)$;

The threshold value γ is no longer needed. From the rules (P1), (B1), and (N1), the (α, β) -probabilistic positive, negative and boundary regions are given, respectively, by:

$$POS_{(\alpha,\beta)}(C) = \{x \in U \mid \Pr(C \mid [x]) \geq \alpha\},$$
$$BND_{(\alpha,\beta)}(C) = \{x \in U \mid \beta < \Pr(C \mid [x]) < \alpha\},$$
$$NEG_{(\alpha,\beta)}(C) = \{x \in U \mid \Pr(C \mid [x]) \leq \beta\}.$$

They are referred to as the three probabilistic regions (Greco, Matarazzo, & Słowínski, 2009; Herbert & Yao, 2009; Pawlak, Wang, & Ziarko, 1988; Slezak & Ziarko, 2002, 2005; Yao, 2007a; Ziarko, 1993). Therefore, the decision-theoretic rough set model provides both a theoretical basis and a practical interpretation of the probabilistic rough sets. The threshold values can be systematically calculated from a loss function based on the Bayesian decision procedure. Other probabilistic rough set models, such as the 0.5-probabilistic rough sets (Pawlak, Wang, &Ziarko, 1988), and the variable precision rough set model (Ziarko, 1993), can be derived from this approach.

INTERPRETATIONS OF CONCEPTS WITH GRANULAR COMPUTING

Granular computing is an emerging field of study that attempts to formalize and explore methods and heuristics of human problem solving with multiple levels of granularity and abstraction

(Bargiela & Pedrycz, 2002; Yao, 2004a, 2004b, 2007b; Zadeh, 1997). A fundamental issue of granular computing is the representation and utilization of granules and granular structures. In this paper, we explore a connection of granules and concepts (Yao, 2009) in classification and learning. Concepts are assumed to be the basic units of knowledge, which play an important role in the study of psychology, cognitive science, and inductive learning (Mitchell, 1982, 1997; Michalski, Carbonell, & Mitchell, 1983; Smith, 1989; Sowa, 1984; van Mechelen, Hampton, Michalski, & Theuns, 1993; Wang, 2007). Following the classical interpretation of a concept (Demri & Orlowska, 1997; Michalski, Carbonell, & Mitchell, 1983; Wille, 1992), we interpret a granule as a pair of a set of objects and a logic formula describing the granule (Zhou &Yao, 2008). The detailed formulations are introduced as follows.

With respect to a dataset, we can build a model based on an information table, in which a set of objects is described by a set of attributes (Pawlak, 1991):

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Table 1. An information table

Object	Weight	Hair	Eyes	Class
o_1	normal	red	blue	+
o_2	Low	dark	brown	+
o_3	low	grey	blue	+
o_4	high	red	blue	+
o_5	low	blond	brown	-
o_6	high	dark	blue	-
o_7	low	red	brown	+
o_8	low	blond	blue	+
o_9	low	grey	brown	-
o_{10}	normal	dark	brown	+
o_{11}	high	dark	brown	-

$$S = (U, At, \{Va \mid a \in At\}, \{Ia \mid a \in At\}),$$

where U is a finite nonempty set of objects, At is a finite nonempty set of attributes, Va is a nonempty set of values of $a \in At$, and $Ia : U \rightarrow Va$ is an information function that maps an object in U to exactly one value in Va. In classification problems, we consider an information table of the forms $S = (U, At = A \cup \{D\}, \{Va\}, \{Ia\})$, where A is a set of condition attributes describing the objects, and D is a decision attribute that indicates the classes of objects.

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

$$IND(A) = \{(x, y) \in U \times U \mid \forall a \in A(Ia(x) = Ia(y))\}.$$

Two objects in U satisfy $IND(A)$ if and only if they have the same values on all attributes in A. The relation $IND(A)$ is called A-indiscernibility relation. The partition of U is a family of all equivalence classes of $IND(A)$ and is denoted by $U/IND(A)$ (U/A). The equivalence classes of the A-indiscernibility relation are denoted as $[x]_A$. Different attribute subsets will give different equivalence classes. For example, Table 1 is a simple information table. The column labeled by Class denotes an expert's classification of the objects. In Table 1, if attribute $A = \{Eyes\}$ is chosen, we can obtain the following family of equivalence classes, or a partition of U:

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Figure 1. General schema of classification process in search of effective granulization

Classify (examples)

```
Use the entire set U as the unlabeled root node of a decision tree;
While there is an unlabeled leaf node in the tree
    Choose an unlabeled leaf node;
    If the conditional probability is above or equal to
        Then change the node to a labeled leaf node with label = "accept";
    Else if the conditional probability is below or equal to
        Then change the node to a labeled leaf node with label = "reject";
    Else if the granule meets certain conditions
        Then change the node to a labeled leaf node with label = "deferment";
    Else replace the unlabeled node with an attribute with each branch
        corresponds to an attribute value, divide the granule into unlabeled
        nonempty sub-granules based on the attribute value
End
Return a ternary classification tree.
```

$$[x]_{\{Eyes\}} = \{\{o_1, o_3, o_4, o_6, o_8\}, \{o_2, o_5, o_7, o_9, o_{10}, o_{11}\}\}.$$

If we consider attribute $A = \{Eyes, Weight\}$, the family of equivalence classes is:

$$[x] = \{\{o_1\}, \{o_2, o_5, o_7, o_9\}, \{o_3, o_8\}, \{o_4, o_6\}, \{o_{10}\}, \{o_{11}\}\}.$$

If we consider each equivalence class as a granule, by choosing different set of attributes from an information table, different granularity can be produced. For certain applications, we may only need to look at granularity of certain level.

Traditionally, a concept is interpreted as a pair of intension and extension. The intension of a concept is given by a set of properties. In order to formally define intensions of concepts, we adopt the decision logic language *L* used and studied by Pawlak (2010). Formulas of *L* are constructed recursively based on a set of atomic formulas corresponding to some basic concepts. An atomic formula is given by $a = v$, where $a \in At$ and $v \in Va$. For each atomic formula $a = v$, an object x satisfies it if $Ia(x) = v$. Otherwise, it does not satisfy $a = v$. From atomic formulas, we can construct other formulas by applying the logic connectives $\neg, \wedge, \vee, \rightarrow,$ and $\leftrightarrow$. Each formula represents an intension of a concept. For two formulas ϕ and φ, we say that ϕ is more specific than φ, and φ is more general than ϕ, if and only if $\phi \rightarrow \varphi$, namely, φ logically follows from ϕ. In other words, the formula $\phi \rightarrow \varphi$ is satisfied by all objects with respect to any universe *U* and any information function *Ia*. If ϕ is more specific than φ, we write $\phi \prec \varphi$, and call ϕ a sub-concept of φ, and φ a super-concept of ϕ.

In inductive learning and concept formation, extensions of concepts are normally defined with respect to a particular training set of examples. If ϕ is a formula, the set $m(\phi)$ is called the meaning of the formula ϕ in *M*. The meaning of a formula ϕ is therefore the set of all objects having the property expressed by the formula ϕ. In other words, ϕ can be viewed as the description of the set of objects m(ϕ). Thus, a connection between formulas and subsets of

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

U is established. For example, in Table 1, if attribute $A = \{Eyes\}$ is chosen, the two equivalence classes can be written as:

$$m(Eyes = blue) = \{o_1, o_3, o_4, o_6, o_8\},$$

$$m(Eyes = brown) = \{o_2, o_5, o_7, o_9, o_{10}, o_{11}\},$$

where Eyes=blue and Eyes=brown are the intensions of the concepts described by the formulas of the language L.

With the introduction of language L, we have a formal description of concepts. A concept definable in a model S is a pair (ϕ ,m(ϕ)), where $\phi \in L$. More specifically, ϕ is a description of m(ϕ) in S, the intension of concept (ϕ ,m(ϕ)), and m(ϕ) is the set of objects satisfying ϕ, the extension of concept (ϕ ,m(ϕ)). A concept (ϕ ,m(ϕ)) is said to be a sub-concept of another concept (ϕ ,m(ϕ)), or (ϕ ,m(ϕ)) a super-concept of (ϕ ,m(ϕ)), in an information table if m(ϕ) $\subseteq$ m(φ). A concept (ϕ ,m(ϕ)) is said to be a smallest non-empty concept in M if there does not exist another non-empty proper subconcept of (ϕ ,m(ϕ)).

Concept learning, to a large extent, depends on the structures of concept space and the target concepts. In general, one may not be able to obtain an effective and efficient learning algorithm, if no restrictions are imposed on the concept space. For this reason, each learning algorithm typically focuses on a specific type of concept.

AN ADAPTIVE LEARNING ALGORITHM FOR CLASSIFICATION

For classification problems, ID3 (Quinlan, 1983) is a well-known algorithm used to generate a decision tree by sequentially choosing the attribute that gives the most information about the class label, the leaf node is added to a branch if the subset of examples for that branch have the same classification labels (i.e., Pr(C|[x]) = 1 or Pr(C|[x]) = 0). In this paper, we introduce an adaptive learning method to construct a decision tree for classification based on three-way decisions with DTRS (Yao, 2010). A subtree is added recursively if the conditional probability lies in between of the two threshold values α and β Otherwise, a leaf node is added if the conditional probability Pr(C|[x]) $\geq \alpha$, or Pr(C|[x]) $\leq \beta$.

In Search of Effective Granulization

Instead of generating the decision tree based on the information gain, in our approach, the attributes of each inner node of the decision tree are sequentially selected by searching for the most suitable granulization at each level. More specifically, we start from the bigger granule at the top level, if the classification decisions cannot be made based on this granulization, we then search for the smaller granules by adding more attribute as inner nodes, until all the examples are correctly classified or certain condition is met. If the granulization at the current level is sufficient for classification, a finner granulization may not be needed at all; this ensures the generated decision tree to be "almost minimal." Based on these principles, the general scheme of the classification process is shown in Figure 1.

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

An Adaptive Learning Algorithm

The actual learning process of building the decision tree is a little bit more complicated. There are a few important steps involved in the learning method, which can be described as follows.

Step 1, At the top most level, the attribute has the least attribute values are selected as the root node A, which will give us the largest granulation. If there is more than one attribute satisfies this condition, we choose the attribute that has the minimum number of objects in its deferment area.

Step 2, A new branch is added for each possible value v_i of A. Estimate the conditional probability of each branch with respect to A, where Pr(C|[x])can be estimated from the frequencies of the training data by putting:

$$\Pr(C \mid [x]) = \frac{|\, C \cap [x] \,|}{|\, [x] \,|},$$

where [x] = m(v_i). If Pr(C|[x]) $\geq \alpha$, objects of this branch belong to the positive region, add a leaf node labeled as C; if Pr(C|[x]) $\leq \beta$, objects of this branch belong to the negative region, add a leaf node labeled as C^C . Otherwise, the classification of this branch cannot be determined, we then search for the next suitable granularity.

Step 3, At the next level, choose an attribute A_j that has the least attribute values from (Attributes – {A}) as the child node for the branch. This time, [x] = m($v_i \wedge v_j$), where v_j represents the possible values of A_j . Similarly, if there is more than one attribute satisfies this condition, we choose the attribute that has the minimum number of objects in its deferment area. Repeat Step 2. Until we can find a leaf node for each branch. Figure 2 shows such an algorithm.

AN ILLUSTRATIVE EXAMPLE

We illustrate the classification process introduced in the previous section by using an information table as Table 1.

At the top level, attribute Eyes is chosen as the root node since it has the least attribute values. Two branches are added corresponding to two possible value Eyes=blue and Eyes=brown, which divides the data set into two granules:

$$m(Eyes = blue) = \{o_1, o_3, o_4, o_6, o_8\},$$
$$m(Eyes = brown) = \{o_2, o_5, o_7, o_9, o_{10}, o_{11}\}.$$

The conditional probability of each equivalence class can be calculated as follows:

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Figure 2. An adaptive learning algorithm in search for effective granulization

```
ALM (S, Decision Attribute, Attributes)
Input:      a training set of examples S,
Output:     a decision tree that correctly classifies all examples in S.
Procedure: Create a root node for the tree;
  If all examples are positive, Return the single-node tree
     Root, with label = +;
    If all examples are negative, Return the single-node tree
     Root, with label = -;
    If number of condition attributes is empty, then Return the
     single node tree Root, with label = most common value of
     the decision attribute in the examples;
Otherwise Begin
  A =The Attribute that has the least number of attribute
   values;
  Decision Tree attribute for Root = A;
  For each possible value, v_i  of A
   Add a new tree branch below Root, corresponding to the test
    A = v_i ;
   Estimate Pr(C|[x]) using Pr(C|[x]) = Pr(C|[x]) = |C ∩ [x]| / |[x]|;

   Let S(v_i), be the subset of examples that have the value  v_i
    for A;
   If Pr(C|[x]) ≥ α
              Then below this new branch add a leaf node with
               label = +;
   Else if Pr(C|[x]) ≤ β
              Then below this new branch add a leaf node with
               label = -;
   Else if S(v_i) is empty
              Then below this new branch add a leaf node with
               label = most common class value in the examples;
   Else below this new branch add the subtree ALM (S(v_i),
    Decision Attribute, Attributes -{A};
End
Return Root
```

$$\Pr(C \mid [o_1]_{\{Eyes\}}) = \frac{|\{o_{1,3,4,6,8}\} \cap \{o_{1,2,3,4,7,8,10}\}|}{|\{o_{1,3,4,6,8}\}|} = \frac{4}{5},$$

$$\Pr(C \mid [o_2]_{\{Eyes\}}) = \frac{|\{o_{2,5,7,9,10,11}\} \cap \{o_{1,2,3,4,7,8,10}\}|}{|\{o_{2,5,7,9,10,11}\}|} = \frac{1}{2}.$$

Assume the two threshold parameters calculated from the loss functions are $\alpha = 0.6$ and $\beta = 0.4$. We have $\Pr(C \mid [o_1]_{\{Eyes\}}) \geq \alpha$, objects $\{o_1, o_3, o_4, o_6, o_8\}$ belong to the positive region, a leaf node can be added to this branch with the class label = +. We also have $\beta \leq \Pr(C \mid [o_2]_{\{Eyes\}})$

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

$\leq \alpha$, objects $\{o_2, o_5, o_7, o_9, o_{10}, o_{11}\}$ belong to the boundary region and need to be further analyzed.

At the second level, attribute Weight is chosen since it has less attribute values than attribute Hair. A subtree is added to the Eyes=brown branch, with three new branches corresponding to three possible values, that is, Weight=normal, Weight=high, and Weight=low, which divides the data set into three granules:

$$m(Eyes = brown \wedge Weight = normal) = \{o_{10}\},$$
$$m(Eyes = brown \wedge Weight = high) = \{o_{11}\},$$
$$m(Eyes = brown \wedge Weight = low) = \{o_2, o_5, o_7, o_9\}.$$

The conditional probability of each equivalence class can be calculated as follows:

$$\Pr(C \mid [o_{10}]_{\{Eyes,Weight\}}) = \frac{|\{o_{10}\} \cap \{o_{1,2,3,4,7,8,10}\}|}{|\{o_{10}\}|} = 1,$$
$$\Pr(C \mid [o_{11}]_{\{Eyes,Weight\}}) = \frac{|\{o_{11}\} \cap \{o_{1,2,3,4,7,8,10}\}|}{|\{o_{11}\}|} = 0,$$
$$\Pr(C \mid [o_2]_{\{Eyes,Weight\}}) = \frac{|\{o_{2,5,7,9}\} \cap \{o_{1,2,3,4,7,8,10}\}|}{|\{o_{2,5,7,9}\}|} = \frac{1}{2}.$$

Assume that we use the same threshold parameters α and β We have $\Pr(C|[o_{10}]_{\{Eyes,Weight\}}) \geq \alpha$, objects $\{o_{10}\}$ belong to the positive region, a leaf node can be added to this branch with the class label = +. We also have $\Pr(C|[o_{10}]_{\{Eyes,Weight\}}) \leq \beta$, objects $\{o_{11}\}$ belong to the negative region, a leaf node can be added to this branch with the class label=-. Finally, $\beta \leq \Pr(C \mid [o_2]_{\{Eyes,Weight\}}) \leq \alpha$, objects $\{o_2,o_5,o_7,o_9\}$ belong to the boundary region and need to be further analyzed.

At the third level, attribute Hair is chosen. A subtree is added to the Eyes=brown $\wedge$ Weight=low branch, with four new branches corresponding to four possible values, that is, Hair=red, Hair=blond, Hair=grey and Hair=dark, which divides the data set into four granules:

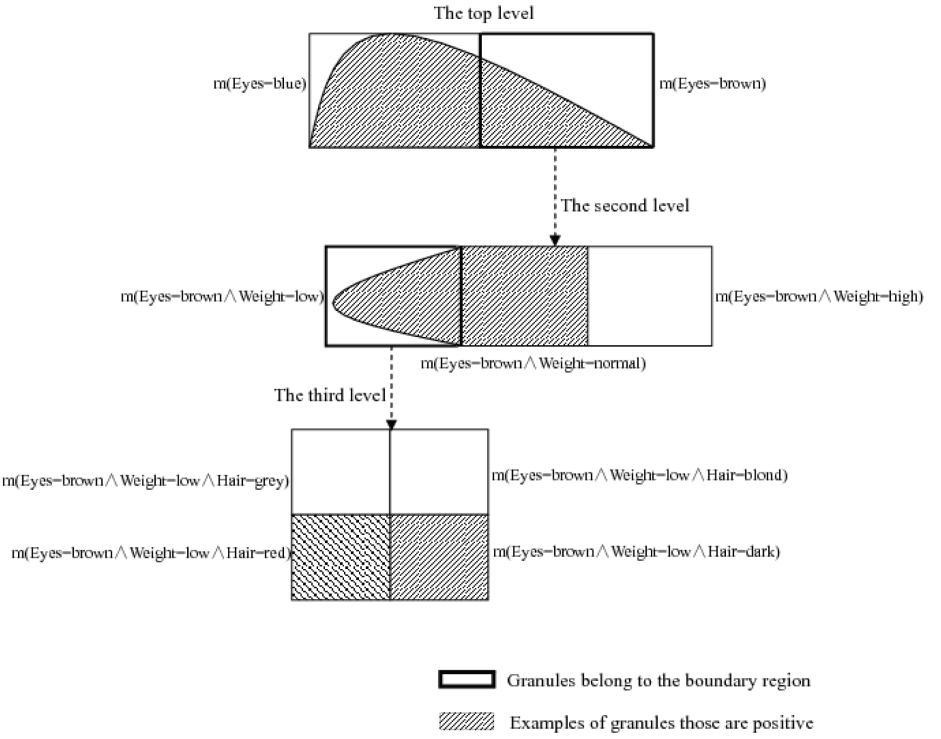
$$m(Eyes = brown \wedge Weight = low \wedge Hair = red) = \{o_7\},$$
$$m(Eyes = brown \wedge Weight = low \wedge Hair = blond) = \{o_5\},$$
$$m(Eyes = brown \wedge Weight = low \wedge Hair = grey) = \{o_9\},$$
$$m(Eyes = brown \wedge Weight = low \wedge Hair = dark) = \{o_2\}.$$

The conditional probability of each equivalence class can be calculated as follows:

$$\Pr(C \mid [o_7]_{\{Eyes,Weight,Hair\}}) = \frac{|\{o_7\} \cap \{o_{1,2,3,4,7,8,10}\}|}{|\{o_7\}|} = 1,$$
$$\Pr(C \mid [o_5]_{\{Eyes,Weight,Hair\}}) = \frac{|\{o_5\} \cap \{o_{1,2,3,4,7,8,10}\}|}{|\{o_5\}|} = 0,$$
$$\Pr(C \mid [o_9]_{\{Eyes,Weight,Hair\}}) = \frac{|\{o_9\} \cap \{o_{1,2,3,4,7,8,10}\}|}{|\{o_9\}|} = 0,$$
$$\Pr(C \mid [o_2]_{\{Eyes,Weight,Hair\}}) = \frac{|\{o_2\} \cap \{o_{1,2,3,4,7,8,10}\}|}{|\{o_2\}|} = 1.$$

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Figure 3. The classification process in search of effective granulization based on three-way decisions



Assume that we use the same threshold parameters α and β. We have $\{o_7\}$ and $\{o_2\}$ in the positive region, a leaf node can be added to both of these branches with class label = +. We also have $\{o_9\}$ and $\{o_5\}$ in the negative region, a leaf node can be added to both of these branches with class label = -. Up to this point, all the objects in the data set have been classified, the decision tree is complete. This classification process in search for effective granulization based on three-way decisions is illustrated in Figure 3.

CONCLUSION AND FUTURE WORK

Decision-theoretic rough set model provides a ternary classification method for classification problems. The deferred examples must be reexamined by collecting further information. We argue that this process can be done automatically by searching for effective granulization. An adaptive learning algorithm is proposed for this purpose, which generates a decision tree by sequentially choosing the attributes that give the most appropriated granulization. We start from the bigger granule at the top level of the tree, if the classification decisions can be made based on this granularity, a finer granularity may not be needed at all, and this ensures the generated tree to be "almost minimal." An illustrative example is given at the end of the paper to demonstrate this process. For future work, we will test our proposed method in larger scale real world data

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

sets, and compare our results with existing classification algorithms to verify the effectiveness of our method.

ACKNOWLEDGMENT

The first author is supported by an NSERC Alexander Graham Bell Canada Graduate Scholarship. The second author is partially supported by an NSERC Canada Discovery grant.

REFERENCES

Bargiela, A., & Pedrycz, W. (2002). *Granular computing: An introduction*. Boston, MA: Kluwer Academic.

Demri, S., & Orlowska, E. (1997). Logical analysis of indiscernibility. In Orlowska, E. (Ed.), *Incomplete information: Rough set analysis* (pp. 347–380). Heidelberg, Germany: Physica Verlag.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York, NY: John Wiley & Sons.

Greco, S., Matarazzo, B., & Słowínski, R. (2009). Parameterized rough set model using rough membership and Bayesian confirmation measures. *International Journal of Approximate Reasoning*, *49*, 285–300. doi:10.1016/j.ijar.2007.05.018

Herbert, J. P., & Yao, J. T. (2009). Game-theoretic rough sets. *Fundamenta Informaticae*, *108*(3-4), 267–286.

Liang, J. Y., & Qian, Y. H. (2008). Information granules and entropy theory in information systems. *Science in China F, 51*(9).

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (1983). *Machine learning, an artificial intelligence approach*. San Francisco, CA: Morgan Kaufmann.

Mitchell, T. M. (1982). Generalization as search. *Artificial Intelligence*, *18*, 203–226. doi:10.1016/0004-3702(82)90040-6

Mitchell, T. M. (1997). *Machine learning*. New York, NY: McGraw-Hill.

Pawlak, Z. (1991). *Rough sets: Theoretical aspects of reasoning about data*. Boston, MA: Kluwer Academic.

Pawlak, Z., Wong, S. K. M., & Ziarko, W. (1988). Rough sets: probabilistic versus deterministic approach. *International Journal of Man-Machine Studies*, *29*, 81–95. doi:10.1016/S0020-7373(88)80032-4

Qian, Y. H., Liang, J. Y., & Dang, C. Y. (2009). Knowledge structure, knowledge granulation and knowledge distance in a knowledge base. *International Journal of Approximate Reasoning*, *50*(1), 174–188. doi:10.1016/j.ijar.2008.08.004

Quinlan, J. R. (1983). Learning efficient classification procedures and their application to chess endgames. In Michalski, J. S., Carbonell, J. G., & Michell, T. M. (Eds.), *Machine learning: An artificial intelligence approach* (*Vol. 1*, pp. 463–482). San Francisco, CA: Morgan Kaufmann.

Slezak, D., & Ziarko, W. (2002, December 9). Bayesian rough set model. In *Proceedings of the Conference on the Foundation of Data Mining*, Maebashi, Japan (pp. 131-135).

Slezak, D., & Ziarko, W. (2005). The investigation of the Bayesian rough set model. *International Journal of Approximate Reasoning*, *40*, 81–91. doi:10.1016/j.ijar.2004.11.004

Smith, E. E. (1989). Concepts and induction. In Posner, M. I. (Ed.), *Foundations of cognitive science* (pp. 501–526). Cambridge, MA: MIT Press.

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Sowa, J. F. (1984). *Conceptual structures, information processing in mind and machine*. Reading, MA: Addison-Wesley.

van Mechelen, I., Hampton, J., Michalski, R. S., & Theuns, P. (Eds.). (1993). *Categories and concepts, theoretical views and inductive data analysis*. New York, NY: Academic Press.

Wang, Y. (2007). Cognitive informatics: exploring theoretical foundations for natural intelligence, neural Informatics, autonomic computing, and agent systems. *International Journal of Cognitive Informatics and Natural Intelligence*, *1*, 1–10. doi:10.4018/jcini.2007040101

Wang, Y., Widrow, B. C., Zhang, B., Kinsner, W., Sugawara, K., & Sun, F. C. (2011). Perspectives on the field of cognitive informatics and its future development. *International Journal of Cognitive Informatics and Natural Intelligence*, *5*(1), 1–17.

Wang, Y., Zadeh, L. A., & Yao, Y. (2009). On the system algebra foundations for granular computing. *International Journal of Software Science and Computational Intelligence*, *1*(1), 64–86. doi:10.4018/jssci.2009010105

Wille, R. (1992). Concept lattices and conceptual knowledge systems. *Computers & Mathematics with Applications (Oxford, England)*, *23*, 493–515. doi:10.1016/0898-1221(92)90120-7

Yao, Y. Y. (2004a). A partition model of granular computing. In J. F. Peters, A. Skowron, J. W. Grzymala-Busse, B. Kostek, R. W. Swiniarski, & M. S. Szczuka (Eds.), *Transactions on Rough Sets I* (LNCS 3100, pp. 232-253).

Yao, Y. Y. (2004b). Granular computing. *Computer Science*, *31*, 1–5.

Yao, Y. Y. (2007a). Decision-theoretic rough set models. In J. T. Yao, P. Lingras, W.-Z. Wu, M. Szczuka, N. J. Cercone, & D. Slezak (Eds.), *Proceedings of the Second International Conference on Rough Sets and Knowledge Technology* (LNCS 4481, pp. 1-12).

Yao, Y. Y. (2007b). The art of granular computing. In M. Kryszkiewicz, J. F. Peters, H. Rybinski, & A. Skowron (Eds.), *Proceeding of the International Conference on Rough Sets and Emerging Intelligent Systems Paradigms* (LNCS 4585, pp. 101-112).

Yao, Y. Y. (2009). Interpreting concept learning in cognitive informatics and granular computing. *IEEE Transactions on Systems, Man, and Cybernetics: Part B, Cybernetics*, *39*(4), 855–866. doi:10.1109/TSMCB.2009.2013334

Yao, Y. Y. (2010). Three-way decisions with probabilistic rough sets. *Information Sciences*, *180*(3), 341–353. doi:10.1016/j.ins.2009.09.021

Yao, Y. Y., & Wong, S. K. M. (1992). A decision theoretic framework for approximating concepts. *International Journal of Man-Machine Studies*, *37*, 793–809. doi:10.1016/0020-7373(92)90069-W

Yao, Y. Y., Wong, S. K. M., & Lingras, P. (1990). A decisiontheoretic rough set model. In Ras, Z. W., Zemankova, M., & Emrich, M. L. (Eds.), *Methodologies for intelligent systems* (*Vol. 5*, pp. 17–24). New York, NY: North-Holland.

Zadeh, L. A. (1997). Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, *19*, 111–127. doi:10.1016/S0165-0114(97)00077-8

Zhou, B., & Yao, Y. Y. (2008). A logic approach to granular computing. *International Journal of Cognitive Informatics and Natural Intelligence*, *2*(2), 63–79. doi:10.4018/jcini.2008040104

Ziarko, W. (1993). Variable precision rough sets model. *Journal of Computer and System Sciences*, *46*, 39–59. doi:10.1016/0022-0000(93)90048-2

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Bing Zhou is a PhD candidate in the department of computer science, University of Regina, Canada. She received her MS in computer science from the University of Regina in 2008. Her research interests include machine learning, data mining, granular computing and formal languages.

Yiyu Yao is a professor with the department of computer science, University of Regina, Canada. He received his B.Eng (1983) in computer science from Xi'an Jiaotong University, China, MS (1988) and PhD (1991) in computer science from University of Regina, Canada. From 1992 to 1998, he was an Assistant Professor and an Associate Professor with the Department of Mathematical Science, Lakehead University, Canada. Dr. Yao's research interests include Web Intelligence, information retrieval, uncertainty management (fuzzy sets, rough sets, interval computing, and granular computing), data mining, and intelligent information systems. He was invited to give talks at many international conferences and universities. Dr. Yao is the chair of steering committee of the International Rough Set Society, a member Technical Committee of Web Intelligence Consortium. He has served, and is serving, as Conference Chair, Program Committee Chair, member of Advisory and/or Program Committee of many international conferences. He is serving as an editorial board member of several international journals.

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.