Dynamic Reduct from Partially Uncertain Data Using Rough Sets

Salsabil Trabelsi[1], Zied Elouedi[1], and Pawan Lingras[2]

[1] Larodec, Institut Superier de Gestion de Tunis, Tunisia
[2] Saint Mary's University Halifax, Canada

Abstract. In this paper, we deal with the problem of attribute selection from a sample of partially uncertain data. The uncertainty exists in decision attributes and is represented by the Transferable Belief Model (TBM), one interpretation of the belief function theory. To solve this problem, we propose dynamic reduct for attribute selection to extract more relevant and stable features for classification. The reduction of the uncertain decision table using this approach yields simplified and more significant belief decision rules for unseen objects.

Keywords: Rough sets, belief function theory, uncertainty, dynamic reduct.

1 Introduction

Feature selection is an important pre-processing stage in machine learning. Rough set theory provides an attractive mechanism for feature selection [5,6,8]. The simplest approach is based on the calculation of reduct. Another issue in real world applications is the uncertain, imprecise or incomplete data. Many researches have adapted rough sets to such an uncertain environment. These extensions do not deal with partially uncertain decision attribute values. In this paper, we deal with the problem of attribute selection from partially uncertain data based on rough sets. The uncertainty exists in decision attributes and is represented by the Transferable Belief Model (TBM), one interpretation of the belief function theory. However, computing reducts from uncertain and noisy data make the results unstable, and sensitive to the sample data. All of these limit the application of rough set theory. Dynamic reducts [1] can lead to better performance in very large datasets, and also provide the ability to accommodate noisy data. The rules calculated by means of dynamic reducts are better predisposed to classify unseen cases, because these reducts are in some sense the most stable reducts, and they appear most frequently in sub-decision systems created by random samples of a given decision system. This paper is organized as follows: Section 2 provides an overview of the rough set theory. Section 3 introduces the belief function theory as understood in the TBM. In Section 4, we propose a new approach to feature selection based on dynamic reducts from partially uncertain data.

H. Sakai et al. (Eds.): RSFDGrC 2009, LNAI 5908, pp. 160–167, 2009.

2 Rough Sets

The idea of rough sets was introduced by Pawlak [6] to deal with imprecise and vague concepts. Here, we introduce only the basic notations. A decision table is an information system of the form $A =$ (U, C $\cup$ {d}), where $d \notin C$ is a distinguished attribute called *decision*. The value set of d is called $\Theta = \{d_1, d_2, ...d_s\}$. In this paper, the notation $c_i(o_j)$ is used to represent the value of a condition attribute $c_i \in C$ for an object $o_j \in U$. Similarly, the notation $d(o_j)$ represents the value of the decision attribute d for an object o_j. The rough sets adopt the concept of indiscernibility relation [6] to partition the object set U into disjoint subsets, denoted by U/B or IND_B. The partition that includes o_j is denoted $[o_j]_B$.

$$IND_B = U/B = \{[o_j]_B | o_j \in U\} \tag{1}$$

Where

$$[o_j]_B = \{o_i | \forall c \in B\, c(o_i) = c(o_j)\} \tag{2}$$

The equivalence classes based on the decision attribute are denoted by $U/\{d\}$.

$$IND_{\{d\}} = U/\{d\} = \{[o_j]_{\{d\}} | o_j \in U\} \tag{3}$$

Let $B \subseteq C$ and $X \subseteq U$. We can approximate X using only the information contained in B by constructing the $B-$*lower* and $B-$*upper approximations* of X, denoted $\underline{B}$(X) and $\bar{B}$(X), respectively, where

$$\underline{B}(X) = \{o_j | [o_j]_B \subseteq X\} \; and \; \bar{B}(X) = \{o_j | [o_j]_B \cap X \neq \emptyset\} \tag{4}$$

The objects in $\underline{B}(X)$ can be classified with certainty as members of X on the basis of knowledge in B, while the objects in $\bar{B}(X)$ can be only classified as possible members of X on the basis of knowledge in B. $Pos_C(\{d\})$, called a positive region of the partition $U/\{d\}$ with respect to C, is the set of all elements of U that can be uniquely classified to blocks of the partition $U/\{d\}$.

$$Pos_C(\{d\}) = \bigcup_{X \in U/\{d\}} \underline{C}(X) \tag{5}$$

A reduct is a minimal subset of attributes from C that preserves the positive region and the ability to perform classifications as the entire attributes set C. A subset $B \subseteq C$ is a reduct of C with respect to d, iff B is minimal and:

$$Pos_B(\{d\}) = Pos_C(\{d\}) \tag{6}$$

The core is the most important subset of attributes, it is included in every reduct.

$$Core(A,\, d) = \bigcap RED(A,\, d) \tag{7}$$

Where $RED(A, d)$ is the set of all reducts of A relative to d.

If $A = (U, C \cup \{d\})$ is a decision table, then any system $B = (U', C \cup \{d\})$ such that $U' \subseteq U$ is called a subtable of A. Let F be a family of subtables of A [1].

$$DR(A,\, F) = RED(A, d) \cap \bigcap_{B \in F} RED(B,\, d) \tag{8}$$

Any element of $DR(A, F)$ is called an F-dynamic reduct of A. From the definition of dynamic reducts, it follows that a relative reduct of A is dynamic if it is also a reduct of all subtables from a given family F. This notation can be sometimes too restrictive so we apply a more general notion of dynamic reduct. They are called (F, ε)-dynamic reducts, where $1 \geq \varepsilon \geq 0$. The set $DR_\varepsilon(A, F)$ of all (F, ε)-dynamic reducts is defined by

$$DR_\varepsilon(A,\, F) = \left\{ R \in RED(A,\, d) : \frac{|\{B \in F : R \in RED(B,\, d)\}|}{|F|} \geq 1 - \varepsilon \right\} \tag{9}$$

3 Belief Function Theory

In this section, we briefly review the main concepts underlying the belief function theory as interpreted in the Transferable Belief Model (TBM) [9,10]. Let Θ be a finite set of elementary events to a given problem, called the frame of discernment. All the subsets of Θ belong to the power set of Θ, denoted by 2^Θ. The impact of a piece of evidence on the subsets of the frame of discernment Θ is represented by a basic belief assignment (bba). The bba is a function $m : 2^\Theta \rightarrow [0, 1]$ such that:

$$\sum_{E \subseteq \Theta} m(E) = 1 \tag{10}$$

The value *m(E)*, called a basic belief mass (bbm), represents the portion of belief committed exactly to the event E. The bba's induced from distinct pieces of evidence are combined by the rule of combination [11].

$$(m_1 \bigcirc\!\!\!\!\cap\, m_2)(E) = \sum_{F, G \subseteq \Theta : F \cap G = E} m_1(F) \times m_2(G) \tag{11}$$

In the TBM, beliefs to make decisions can be represented by probability functions called the pignistic probabilities denoted $BetP$ and are defined as [10]:

$$BetP(\{a\}) = \sum_{F \subseteq \Theta} \frac{|\ \{a\} \cap F\ |}{|\ F\ |} \frac{m(F)}{(1 - m(\emptyset))}\ , \text{ for all }\ a \in \Theta \tag{12}$$

4 Dynamic Reduct under Uncertainty

Our decision system is characterized by high level of uncertain and noisy data. One of the issues with such a data is that the resulting reducts are not stable, and are sensitive to sampling. The belief decision rules generated are not suitable for classification. The solution to this problem is to redefine the concept of dynamic reduct in the new context as we have done in this paper. The rules calculated by means of dynamic reducts are better predisposed to classify unseen objects, because they are the most frequently appearing reducts in sub-decision systems created by random samples of a given decision system. In this section, we will adapt the basic concepts of rough sets such as decision system, indiscernibility relation, set approximation and positive region in order to redefine the concept of dynamic reduct in the uncertain context. The objective is to extract more stable reducts from the uncertain decision system.

4.1 Basic Concepts of Rough Sets under Uncertainty

Uncertain Decision System. Our uncertain decision system is given by $A = (U, C \cup \{ud\})$, where $U = \{o_j : 1 \leq j \leq n\}$ is characterized by a set of certain condition attributes C=$\{c_1,\ c_2,...,c_k\}$, and an uncertain decision attribute ud. We represent the uncertainty of each object o_j by a bba m_j expressing beliefs on decisions defined on the frame of discernment Θ=$\{ud_1,\ ud_2,...,ud_s\}$ representing the possible values of ud. These bba's are given by an expert.

Example. Let us use Table 1 to describe our uncertain decision system. It contains eight objects, three certain condition attributes C=$\{a, b, c\}$ and an uncertain decision attribute $ud = e$ with two possible values $\{e_1,\ e_2\}$ representing Θ.

Table 1. Uncertain decision table

U	a	b	c	e
o_1	0	0	0	$m_1(\{e_1\}) = 0.95 \quad m_1(\{e_2\}) = 0.05$
o_2	0	1	1	$m_2(\{e_2\}) = 1$
o_3	0	0	2	$m_3(\{e_1\}) = 0.5 \quad m_3(\Theta) = 0.5$
o_4	1	0	2	$m_4(\{e_2\}) = 0.6 \quad m_4(\Theta) = 0.4$
o_5	1	0	2	$m_5(\{e_2\}) = 1$
o_6	0	1	1	$m_6(\{e_2\}) = 0.9 \quad m_6(\{\Theta\}) = 0.1$
o_7	1	0	0	$m_7(\{e_1\}) = 1$
o_8	1	0	1	$m_8(\{e_1\}) = 0.9 \ m_8(\{\Theta\}) = 0.1$

For the object o_3, 0.5 of beliefs are exactly committed to the decision e_1, whereas 0.5 of beliefs is assigned to the whole of frame of discernment Θ (ignorance). With bba, we can represent the certain case, like for the objects o_2, o_5 and o_7. Besides, we can represent probability case, like the bba relative to the object o_1 and possiblitic case like the consonant bba relative to the object o_3. The decision rules induced from the partially uncertain decision system are denoted belief decision rules where the decision is represented by a bba: *If a=0 and b=0 and c=2 Then* $m_3(\{e_1\}) = 0.5 \quad m_3(\Theta) = 0.5$.

Indiscernibility Relation. For the condition attributes, the indiscernibility relation U/C is the same as in the certain case because their values are certain. However, the indiscernibility relation for the decision attribute $U/\{ud\}$ is not the same as in the certain case. The decision value is represented by a bba. So, we need to assign each object to the right equivalence class. The idea is to use the distance between two bba's. Many distance measures between two bba's were developed [2,3,4]. We will choose the distance measure described in [2] which satisfies properties such as non-negativity, non-degeneracy and symmetry.

For every ud_i, an uncertain decision value, we define:

$$X_i = \{o_j | dist(m(ud_i) = 1, m_j) \neq 1\} \tag{13}$$

$$IND_{\{ud\}} = U/\{ud\} = \{X_i | ud_i \in \Theta\} \tag{14}$$

Where $dist$ is a distance measure between two bba's.

$$dist(m_1, m_2) = \sqrt{\frac{1}{2}(\| \vec{m_1} \|^2 + \| \vec{m_2} \|^2 - 2 < \vec{m_1}, \vec{m_2} >)} \tag{15}$$

Where $< \vec{m_1}, \vec{m_2} >$ is the scalar product defined by:

$$< \vec{m_1}, \vec{m_2} >= \sum_{i=1}^{|2^\Theta|} \sum_{j=1}^{|2^N|} m_1(A_i) m_2(A_j) \frac{|A_i \cap A_j|}{|A_i \cup A_j|} \tag{16}$$

with A_i, $A_j \in 2^\Theta$ for $i, j = 1, 2, \cdots, |2^\Theta|$. $\| \vec{m_1} \|^2$ is then the square norm of $\vec{m_1}$.

Example. Let us continue with the same example to compute the equivalence classes based on condition attributes in the same manner as in the certain case: U/C= $\{\{o_1\}, \{o_2, o_6\}, \{o_3\}, \{o_4, o_5\}, \{o_7\}, \{o_8\}\}$ and to compute the equivalence classes based on uncertain decision attribute $U/\{ud\}$ as follows:

For the uncertain decision value $ud_1 = e_1$,

$dist(m(e_1) = 1, m_1) \neq 1$
$dist(m(e_1) = 1, m_2) = 1$
$dist(m(e_1) = 1, m_3) \neq 1$
$dist(m(e_1) = 1, m_4) \neq 1$
$dist(m(e_1) = 1, m_5) \neq 1$
$dist(m(e_1) = 1, m_6) \neq 1$
$dist(m(e_1) = 1, m_7) \neq 1$
$dist(m(e_1) = 1, m_8) \neq 1$.

So, X_1 =$\{o_1$, o_3, o_4, o_5, o_6, o_7, $o_8\}$.

For the uncertain decision value $ud_2 = e_2$,

$dist(m(e_2) = 1, m_1) \neq 1$
$dist(m(e_2) = 1, m_2) \neq 1$
$dist(m(e_2) = 1, m_3) \neq 1$
$dist(m(e_2) = 1, m_4) \neq 1$
$dist(m(e_2) = 1, m_5) \neq 1$
$dist(m(e_2) = 1, m_6) \neq 1$

$dist(m(e_2) = 1, m_7) = 1$
$dist(m(e_2) = 1, m_8) \neq 1$.

So, X_2 =\{o_1, o_2, o_3, o_4, o_5, o_6, o_8\}.
$U/\{ud\}$=\{\{o_1, o_3, o_4, o_5, o_6, o_7, o_8\}, \{o_1, o_2, o_3, o_4, o_5, o_6, o_8\}\}.

Set Approximation. To compute the new *lower* and *upper* approximations for our uncertain decision table, we follow two steps:

1. For each equivalence class from U/C based on condition attributes C, combine their bba using the operator mean. The operator mean is more suitable in our case to combine these bba's than the rule of combination in eq. 11 which is proposed especially to combine different beliefs on decision for one object and not different beliefs for different objects.
2. For each equivalence class X_i from $U/\{ud\}$ based on uncertain decision attribute ud_i, we compute the new *lower* and *upper* approximations, as follows:

$$\underline{C}X_i = \{o_j | [o_j]_C \cap X_i \neq \emptyset \, and \, dist(m(ud_i) = 1, m_{[o_j]_C}) \leq threshold\} \quad (17)$$

In the *lower* approximation, we find all equivalence classes (subsets) from U/C included in X_i such that the distance between the combined bba $m_{[o_j]_C}$ and the certain bba $m(ud_i) = 1$ is less than a *threshold*. (In an uncertain context, the *threshold* is needed to give more flexibility to the set approximations). We compute the *upper* approximation in the same manner as in the certain case.

$$\bar{C}X_i = \{o_j | [o_j]_C \cap X_i \neq \emptyset\} \quad (18)$$

Example. We continue with the same example to compute the new *lower* and *upper* approximations. After the first step, we obtain the combined bba for each equivalence class from U/C using operator mean. Table 2 represents the combined bba for the equivalence classes \{o_2, o_6\} and \{o_4, o_5\}.

Table 2. The combined bba for the subsets \{o_2, o_6\} and \{o_4, o_5\}

Object	$m(\{e_1\})$	$m(\{e_2\})$	$m(\Theta)$
o_2	0	1	0
o_6	0	0.9	0.1
$m_{2,6}$	0	0.95	0.05
o_4	0	0.4	0.6
o_5	0	1	0
$m_{4,5}$	0	0.7	0.3

Next, we compute the *lower* and *upper* approximations for each equivalence class $U/\{ud\}$. We will use $threshold = 0.1$.

For the uncertain decision value ud_1=e_1, let X_1 =\{o_1, o_3, o_4, o_5, o_6, o_7, o_8\}. The subsets \{o_1\}, \{o_3\}, \{o_4, o_5\}, \{o_7\} and \{o_8\} are included in X_1. We should check the distance between their bba and the certain bba $m(e_1) = 1$.

$dist(m(e_1) = 1, m_1) < 0.1$
$dist(m(e_1) = 1, m_3) > 0.1$

$dist(m(e_1) = 1, m_{4,5}) > 0.1$
$dist(m(e_1) = 1, m_7) < 0.1$
$dist(m(e_1) = 1, m_8) < 0.1$
$\underline{C}X_1$={o_1, o_7, o_8} and $\bar{C}X_1$={o_1, o_3, o_4, o_5, o_7, o_8}

For uncertain decision value ud_2=e_2, let X_2 ={o_1, o_2, o_3, o_4, o_5, o_6, o_8}. $\underline{C}X_2$={o_2, o_6} and $\bar{C}X_2$={o_2, o_3, o_4, o_5, o_6}

Positive Region. With the new *lower* approximation, we can redefine the positive region:

$$UPos_C(\{ud\}) = \bigcup_{X_i \in U/\{ud\}} \underline{C}X_i \tag{19}$$

Example: Let us continue with the same example, to compute the positive region of *A*. $UPos_C$({ud})={o_1, o_2, o_6, o_7, o_8}

Reduct and Core. Using the new formalism of positive region, we can redefine the reduct of *A* as a minimal set of attributes $B \subseteq C$ such that:

$$UPos_B(\{ud\}) = UPos_C(\{ud\}) \tag{20}$$

$$UCore(A,\ ud) = \bigcap URED(A,\ ud) \tag{21}$$

Where $URED(A,\ ud)$ is the set of all reducts of *A* relative to *ud*.

Example. Using our example, we find that $UPos_{\{a,c\}}(\{ud\})$= $UPos_{\{b,c\}}(\{ud\})$ $= UPos_C(\{ud\})$. So, we have two possible reducts {a,c} and {b,c}. The attribute *c* is the relative core.

4.2 Dynamic Reduct from Uncertain Data

Using the new definition of reduct in our uncertain context, we can redefine the concept of dynamic reduct as follows:

$$UDR(A,\ F) = URED(A, ud) \cap \bigcap_{B \in F} URED(B,\ ud) \tag{22}$$

Where *F* be a family of subtables of *A*. This notation can be sometimes too restrictive so we apply a more general notion of dynamic reduct. They are called (F, ε)-dynamic reducts, where $1 \geq \varepsilon \geq 0$. The set $UDR_\varepsilon(A,\ F)$ of all (F, ε)-dynamic reducts is defined by:

$$UDR_\varepsilon(A,\ F) = \left\{ R \in URED(A,\ ud) : \frac{|\{B \in F : R \in RED(B,\ ud)\}|}{|F|} \geq 1 - \varepsilon \right\} \tag{23}$$

Example. To compute the dynamic reduct of the uncertain decision system *A*. We divide our uncertain decision system into two subtables *B* and B' to obtain a family *F* of sub-decision system. *B* contains the objects o_1, o_2, o_3, o_4 and B' contains the objects o_5, o_6, o_7, o_8. The two subtables have the same reducts as the whole decision system *A*. So, the subsets {a,c} and {b,c} are dynamic reducts relative to the chosen family *F*.

5 Conclusion and Future Work

In this paper, we have adapted the basic concepts of rough sets such as decision system, indiscernibility relation, set approximation and reduct in an uncertain context. We handle uncertainty in decision attributes using the belief function theory. We further propose dynamic reduct to address the problem of unstable reducts in uncertain decision systems. As a future work, we will experiment with many uncertain databases to evaluate the proposed feature selection based on dynamic reducts.

References

1. Bazan, J., Skowron, A., Synak, P.: Dynamic reducts as a tool for extracting laws from decision tables in: methodologies for intelligent system. In: Raś, Z.W., Zemankova, M. (eds.) ISMIS 1994. LNCS (LNAI), vol. 869, pp. 346–355. Springer, Heidelberg (1994)
2. Bosse, E., Jousseleme, A.L., Grenier, D.: A new distance between two bodies of evidence. In: Information Fusion, vol. 2, pp. 91–101 (2001)
3. Elouedi, Z., Mellouli, K., Smets, P.: Assessing sensor reliability for multisensor data fusion within the transferable belief model. IEEE Trans. Syst. Man cyben. 34(1), 782–787 (2004)
4. Fixen, D.: The modified Dempster-Shafer approach to classification. IEEE Trans. Syst. Man Cybern. A 27(1), 96–104 (1997); In: Workshop Notes, Foundations and New Directions of Data Mining, the 3rd International Conference on Data Mining, Melbourne, Florida, vol. 5663
5. Modrzejewski, M.: Feature selection using rough sets theory. In: Proceedings of the 11th International Conference on Machine Learning, pp. 213–226 (1993)
6. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishing, Dordrecht (1991)
7. Shafer, G.: A mathematical theory of evidence. Princeton University Press, Princeton (1976)
8. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Slowinski, R. (ed.) Intelligent Decision Support, pp. 331–362. Kluwer Academic Publishers, Boston (1992)
9. Smets, P.: The transferable belief model. Artificial Intelligence 66(2), 191–236 (1994)
10. Smets, P.: The transferable belief model for quantified belief representation. In: Gabbay, D.M., Smets, P. (eds.) Handbook of defeasible reasoning and uncertainty management systems, vol. 1, pp. 207–301. Kluwer, Dordrecht (1998)
11. Smets, P.: Application of belief transferable belief model to diagnostic problems. International journal of intelligent systems 13(2-3), 127–157 (1998)