An Optimization Viewpoint of Decision-Theoretic Rough Set Model

Xiuyi Jia[1,3], Weiwei Li[2], Lin Shang[3], and Jiajun Chen[3]

[1] School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, China, 210094
jiaxy@nlp.nju.edu.cn

[2] College of Information Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 210016
amy.vivilee@gmail.com

[3] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China, 210093
{shanglin,chenjj}@nju.edu.cn

Abstract. This paper considers an optimization viewpoint of decision-theoretic rough set model. An optimization problem is proposed by considering the minimization of the decision cost. Based on the optimization problem, cost functions and thresholds used in decision-theoretic rough set model can be learned from the given data automatically. An adaptive learning algorithm Alcofa is proposed. Another significant inference drawn from the solution of the optimization problem is a minimum cost based attribute reduction. The attribute reduction can be interpreted as finding the minimal attribute set to make the decision cost minimum. The optimization viewpoint can bring some new insights into the research on decision-theoretic rough set model.

Keywords: Decision-theoretic rough set model, optimization problem, decision cost, attribute reduction.

1 Introduction

As a kind of probabilistic rough set model, decision-theoretic rough set model (DTRS) [4–6] can derive current several probabilistic rough set models when proper cost functions are used, such as 0.5 probabilistic rough set model [2], variable precision rough set model [11] and Bayesian rough set models [3]. We study the decision-theoretic rough set model from an optimization viewpoint in this paper, and present some new results which may increase the understanding of the decision-theoretic rough set model.

Based on Bayesian decision procedure, decision-theoretic rough set model provides systematic methods for deriving the required thresholds on probabilities for defining the three regions: positive region, boundary region and negative region. One object x is classified into a certain region because the cost generated by classifying x into the region is less than the cost generated by classifying x into

J.T. Yao et al. (Eds): RSKT 2011, LNCS 6954, pp. 457–465, 2011.

the other two regions. All decisions are made based on the minimum cost. Decision cost is an important concept in decision-theoretic rough set model. From the optimization view, we construct an optimization problem which is to minimize the decision cost in decision-theoretic rough set model. Through solving the optimization problem, we can learn the thresholds and proper cost functions from given data without any preliminary knowledge, and we also define a new attribute reduction which is based on the minimum cost. The attribute reduction can be interpreted as finding the minimal attribute set to make the whole decision cost minimum, which is more intuitive and reasonable.

Using proper cost functions, we can derive different thresholds and get the corresponding probabilistic rough set models. The cost functions play an important role in decision-theoretic rough set model. In general, the cost functions are given by experts, and little contributions are on learning cost functions from data. Herbert and Yao [1] have proposed an approach for using a game-theoretic learning method to govern the modification of cost functions in order to improve some measures, this method needs users provide some measures first and define an acceptable levels of tolerance to stop the repeating procedure. Compared to their method, our method based on the optimization viewpoint does not need users' participation, it is automatic and easy to implement.

As to the non-monotonic property of the regions in decision-theoretic rough set model, there exist interpretation difficulties in those attribute reductions which are defined on region preservation [10]. The attribute reduction based on minimum decision cost does not concentrate on preserving any region. The goal of the reduction is to help users make better decisions, which means less decision cost.

The rest of the paper is organized as follows. In Section 2, we review the main ideas of decision theoretic rough set model. In Section 3, we give a detailed explanation of the optimization problem, and by solving the problem, we can learn the thresholds and the cost functions from data. An adaptive learning algorithm is proposed. We also define a new attribute reduction and give some remarks on the optimization problem. Section 4 concludes.

2 Basic Notions of Decision-theoretic Rough Set Model

In this section, we present some basic definitions of decision-theoretic rough set model [7].

Definition 1. *A decision table is the following tuple:*

$$S = (U, At = C \cup \{D\}, \{V_a | a \in At\}, \{I_a | a \in At\}), \tag{1}$$

where U is a finite nonempty set of objects, At is a finite nonempty set of attributes, C is a set of condition attributes describing the objects, and D is a decision attribute that indicates the classes of objects. V_a is a nonempty set of values of $a \in At$, and $I_a : U \to V_a$ is an information function that maps an object in U to exactly one value in V_a.

In the decision table, an object x is described by its equivalence class under a set of attributes $A \subseteq At$: $[x]_A = \{y \in U | \forall a \in A(I_a(x) = I_a(y))\}$. Let $\pi_D = \{D_1, D_2, \ldots, D_m\}$ be a partition of the universe U defined by the decision attribute D.

Let $\Omega = \{\omega_1, \ldots, \omega_s\}$ be a finite set of s states and let $\mathcal{A} = \{a_1, \ldots, a_m\}$ be a finite set of m possible actions. Let $\lambda(a_i|\omega_j)$ denote the cost, for taking action a_i when the state is ω_j. Let $p(\omega_j|x)$ be the conditional probability of an object x being in state ω_j, suppose action a_i is taken. The expected cost associated with taking action a_i is given by:

$$\mathcal{R}(a_i|x) = \sum_{j=1}^{s} \lambda(a_i|\omega_j) \cdot p(\omega_j|x). \tag{2}$$

In decision-theoretic rough set model, the set of states $\Omega = \{X, X^c\}$, indicating that an object is in a decision class X and not in X, respectively. The probabilities for these two complement states can be denoted as $p(X|[x]) = \frac{|X \cap [x]|}{|[x]|}$ and $p(X^c|[x] = 1 - p(X|[x]))$. With respect to the three regions, the set of actions with respect to a state is given by $\mathcal{A} = \{a_P, a_B, a_N\}$, where a_P, a_B, and a_N represent the three actions in classifying an object x, namely, deciding $x \in \mathrm{POS}(X)$, deciding $x \in \mathrm{BND}(X)$, and deciding $x \in \mathrm{NEG}(X)$, respectively. Let λ_{PP}, λ_{BP} and λ_{NP} denote the costs incurred for taking actions a_P, a_B and a_N, respectively, when an object belongs to X, and λ_{PN}, λ_{BN} and λ_{NN} denote the costs incurred for taking the same actions when the object does not belong to X.

Given the cost functions, the expected costs associated with taking different actions for objects in $[x]$ can be expressed as:

$$\begin{aligned}
\mathcal{R}_P &= \mathcal{R}(a_P|[x]) = \lambda_{PP} \cdot p(X|[x]) + \lambda_{PN} \cdot p(X^c|[x]),\\
\mathcal{R}_B &= \mathcal{R}(a_B|[x]) = \lambda_{BP} \cdot p(X|[x]) + \lambda_{BN} \cdot p(X^c|[x]),\\
\mathcal{R}_N &= \mathcal{R}(a_N|[x]) = \lambda_{NP} \cdot p(X|[x]) + \lambda_{NN} \cdot p(X^c|[x]).
\end{aligned} \tag{3}$$

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

(P) If $\mathcal{R}_P \leq \mathcal{R}_B$ and $\mathcal{R}_P \leq \mathcal{R}_N$, decide $x \in \mathrm{POS}(X)$;
(B) If $\mathcal{R}_B \leq \mathcal{R}_P$ and $\mathcal{R}_B \leq \mathcal{R}_N$, decide $x \in \mathrm{BND}(X)$;
(N) If $\mathcal{R}_N \leq \mathcal{R}_P$ and $\mathcal{R}_N \leq \mathcal{R}_B$, decide $x \in \mathrm{NEG}(X)$.

Consider a special kind of cost functions with $\lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}$ and $\lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}$, the decision rules can be reexpressed as:

(P) If $p(X|[x]) \geq \alpha$ and $p(X|[x]) \geq \gamma$, decide $x \in \mathrm{POS}(X)$;
(B) If $p(X|[x]) \leq \alpha$ and $p(X|[x]) \geq \beta$, decide $x \in \mathrm{BND}(X)$;
(N) If $p(X|[x]) \leq \beta$ and $p(X|[x]) \leq \gamma$, decide $x \in \mathrm{NEG}(X)$,

where the parameters α, β, and γ are defined as:

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

$$\beta = \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})},$$
$$\gamma = \frac{(\lambda_{PN} - \lambda_{NN})}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}. \tag{4}$$

Each rule is defined by two out of the three parameters. The conditions of rule (B) suggest that $\alpha > \beta$ may be a reasonable constraint; it will ensure a well-defined boundary region. If we obtain the following condition on the cost functions [7]:

$$\frac{(\lambda_{NP} - \lambda_{BP})}{(\lambda_{BN} - \lambda_{NN})} > \frac{(\lambda_{BP} - \lambda_{PP})}{(\lambda_{PN} - \lambda_{BN})}, \tag{5}$$

then $0 \leq \beta < \gamma < \alpha \leq 1$. In this case, after tie-breaking, the following simplified rules are obtained:

(P1) If $p(X|[x]) \geq \alpha$, decide $x \in \mathrm{POS}(X)$;
(B1) If $\beta < p(X|[x]) < \alpha$, decide $x \in$ BND(X);
(N1) If $p(X|[x]) \leq \beta$, decide $x \in \mathrm{NEG}(X)$.

More different conditions on cost functions were discussed in [7, 9]. By using the thresholds, one can divide the universe U into three regions of a decision partition π_D based on (α, β):

$$\mathrm{POS}_{(\alpha,\beta)}(\pi_D|\pi_A) = \{x \in U | p(D_{\max}([x]_A)|[x]_A) \geq \alpha\},$$
$$\mathrm{BND}_{(\alpha,\beta)}(\pi_D|\pi_A) = \{x \in U | \beta < p(D_{\max}([x]_A)|[x]_A) < \alpha\},$$
$$\mathrm{NEG}_{(\alpha,\beta)}(\pi_D|\pi_A) = \{x \in U | p(D_{\max}([x]_A)|[x]_A) \leq \beta\}, \tag{6}$$

where $D_{\max}([x]_A) = \arg\max_{D_i \in \pi_D}\{\frac{|[x]_A \cap D_i|}{|[x]_A|}\}$.

Unlike rules in the classical rough set theory, all three types of rules may be *uncertain*. They represent the levels of tolerance in making incorrect decisions. Each rule brings corresponding cost as to its error rate. Consider the special case where we assuming zero cost for a correct classification, namely, $\lambda_{PP} = \lambda_{NN} = 0$, and let $p = p(D_{\max}([x]_A)|[x]_A)$, the decision costs of all rules are easily defined as [7]:

$$\text{positive rule} : (1-p) \cdot \lambda_{PN},$$
$$\text{boundary rule} : p \cdot \lambda_{BP} + (1-p) \cdot \lambda_{BN},$$
$$\text{negative rule} : p \cdot \lambda_{NP}. \tag{7}$$

For a given decision table, the decision cost of the table can be expressed as:

$$\mathbf{COST} = \sum_{p_i \geq \alpha} (1-p_i) \cdot \lambda_{PN} + \sum_{\beta < p_j < \alpha} (p_j \cdot \lambda_{BP} + (1-p_j) \cdot \lambda_{BN}) + \sum_{p_k \leq \beta} p_k \cdot \lambda_{NP}, \tag{8}$$

where $p_i = p(D_{\max}([x_i]_A)|[x_i]_A)$.

3 An Optimization Problem in Decision-Theoretic Rough Set Model

In section 2, we have given the expression of the cost of a decision table. According to the Bayesian decision principle, it is better to get a smaller value of the cost, so we propose an optimization problem with the objective of minimizing the value of the cost, written as

$$\min \mathbf{COST}. \tag{9}$$

Based on the optimization problem, we can learn the thresholds and proper cost functions from given data without any preliminary knowledge. We also define a new attribute reduction based on the minimum cost, which can be interpreted as finding the minimal attribute set to make the whole decision costs minimum.

From Equation(8), we know that the decision cost can be formulated by the probabilities of given objects and the cost functions, while the cost functions can be presented by the required thresholds. By solving the optimization problem, firstly, we can get the thresholds and derive the cost functions; secondly, we can find the corresponding probabilities which make the decision cost minimum and get the proper attribute reduction from these probabilities.

3.1 Learning Thresholds and Cost Functions from Data

Based on Equation(8) and Equation(9), we can construct an optimization problem:

$$\begin{aligned} \min_{\alpha,\beta,\gamma} \quad & \sum_{p_i \geq \alpha} (1-p_i) \cdot \lambda_{PN} + \sum_{\beta < p_j < \alpha} (p_j \cdot \lambda_{BP} + (1-p_j) \cdot \lambda_{BN}) + \sum_{p_k \leq \beta} p_k \cdot \lambda_{NP} \\ s.t. \quad & 0 < \beta < \gamma < \alpha < 1. \end{aligned} \tag{10}$$

In Equation(4), three thresholds (α, β, γ) are presented by six cost functions. Assume $\lambda_{PP} = \lambda_{NN} = 0$, which means making right decisions does not bring any cost. Now we can present the rest four cost functions by the three thresholds, reversely.

$$\begin{aligned} \lambda_{PN} &= \lambda_{PN}; \\ \lambda_{NP} &= \frac{1-\gamma}{\gamma} \cdot \lambda_{PN}; \\ \lambda_{BN} &= \frac{\beta \cdot (\alpha - \gamma)}{\gamma \cdot (\alpha - \beta)} \cdot \lambda_{PN}; \\ \lambda_{BP} &= \frac{(1-\alpha) \cdot (\gamma - \beta)}{\gamma \cdot (\alpha - \beta)} \cdot \lambda_{PN}. \end{aligned} \tag{11}$$

Considered with Equation(11) and scaled λ_{PN} to 1, the optimization object in Equation(10) can be reexpressed as:

$$\min_{\alpha,\beta,\gamma} \sum_{p_i \geq \alpha} (1-p_i) + \sum_{\beta < p_j < \alpha} (p_j \cdot \frac{(1-\alpha)\cdot(\gamma-\beta)}{\gamma\cdot(\alpha-\beta)} + (1-p_j)\cdot\frac{\beta\cdot(\alpha-\gamma)}{\gamma\cdot(\alpha-\beta)}) + \sum_{p_k \leq \beta} p_k \cdot \frac{1-\gamma}{\gamma}. \quad (12)$$

We can get the three thresholds and all cost functions through solving the optimization problem now. It is not easy to get the optimum result as the search space for α, β, γ is $(0, 1)$. In this paper, we assume the search space is the set of probabilities of all objects, then we can get the optimum result now and we also give an adaptive learning algorithm **Alcofa** as follows.

The basic idea of **Alcofa** is explained as following. The required thresholds are relative to three special objects, then the values of thresholds are equal to the three special objects' probabilities. The goal of this algorithm is to find the corresponding three probability values. Assume the current thresholds (α, β, γ) are learned from objects $\mathcal{X} = \{x_1, \ldots, x_{i-1}\}$ in the training set, for the next object x_i coming from the training set, the probability p_i is added to compute the overall cost $\mathbf{COST}_{\mathcal{X}\cup\{x_i\}}$ based on thresholds (α, β, γ), denoted as $Min_{\mathbf{COST}}$. Then replace the three thresholds by p_i next in turn to find the new minimal overall cost $\mathbf{COST}'_{\mathcal{X}\cup\{x_i\}}$. If $\mathbf{COST}'_{\mathcal{X}\cup\{x_i\}} < Min_{\mathbf{COST}}$, then current thresholds (α, β, γ) are updated to thresholds (α', β', γ'); else, (α, β, γ) are unchanged. Same procedure is applied to deal the next object x_{i+1}, the loop is ended until all objects in the training set are finished. The final thresholds are the result for getting the minimum overall cost.

The key step of the algorithm is the replacement of one threshold by the probability. For example, let $\alpha' = p_i$ now, if $\gamma < p_i$ and $\beta < p_i$, then β, γ are unchanged. If $p_i \leq \beta$ or $p_i \leq \gamma$, it is inconsistent with $\beta < \gamma < \alpha$, then γ and β will change to two smaller corresponding numbers than α, for example, $\gamma' = p_i \cdot (1 - 0.01)$ and $\beta' = p_i \cdot (1 - 0.05)$. The new thresholds are denoted as (α', β', γ').

Table 1 shows the description of the adaptive learning algorithm.

The computational complexity of **Alcofa** is $O(n^2)$. For the probability of object x_i: p_i, we can compute it by using the rough set method or get it from other classifiers, e.g. Naive Bayesian classifier, which makes the algorithm more robust and practical. Yao and Zhou [8] also have proposed a Naive Bayesian rough set model by combing Naive Bayesian classifier with decision-theoretic rough set model.

3.2 Attribute Reduction Based on Minimal Decision Cost

In paper [10], the authors proposed three problems to show the difficulties with the interpretations of those attribute reductions which are defined based on the positive region preservation. An important reason of producing these difficulties is the non-monotonicity property of the regions in decision-theoretic rough set model. The decrease of an attribute can result a decrease or an increase or a constant of a probabilistic region(positive, boundary, or negative region). In this section, we define a new attribute reduction which is based on the above optimization problem solving and irrelevant to those regions.

Table 1. Adaptive Learning COst Functions Algorithm (**Alcofa**)

```
Input:   training set X = {x_1, ..., x_n}.
Output: three thresholds (α, β, γ).
BEGIN
    initialize α, γ, β, 𝒳 = ∅, min=MAXINT;
    FOR each x_i ∈ X
        𝒳 = 𝒳 ∪ {x_i};
        compute the overall cost COST_𝒳 based on (α, β, γ);
        min = COST_𝒳;
        replace α by p_i;
        compute the overall cost COST'_𝒳 based on (α', β, γ);
        IF COST'_𝒳 < min
            min = COST'_𝒳;
        END IF
        replace β by p_i;
        compute the overall cost COST'_𝒳 based on (α, β', γ);
        IF COST'_𝒳 < min
            min = COST'_𝒳;
        END IF
        replace γ by p_i;
        compute the overall cost COST'_𝒳 based on (α, β, γ');
        IF COST'_𝒳 < min
            min = COST'_𝒳;
        END IF
        update the thresholds (α, β, γ) to (α', β', γ')
                  corresponding to the min value;
    END FOR
    return the current thresholds (α, β, γ);
END BEGIN
```

Based on attribute set $A \subseteq C$, we can rewrite the cost formulation as:

$$\mathbf{COST}_A = \sum_{x_i \in \mathrm{POS}_{(\alpha,\beta)}(\pi_D|\pi_A)} (1 - p_i) \cdot \lambda_{PN} + \sum_{p_j \in \mathrm{BND}_{(\alpha,\beta)}(\pi_D|\pi_A)} (p_j \cdot \lambda_{BP} + (1 - p_j) \cdot \lambda_{BN}) + \sum_{p_k \in \mathrm{NEG}_{(\alpha,\beta)}(\pi_D|\pi_A)} p_k \cdot \lambda_{NP}. \quad (13)$$

Then the optimization problem is described as finding proper attribute set to make the whole decision cost minimum. Based on that, an attribute reduction definition based on minimum decision cost is given as following:

Definition 2. *In a decision table* $S = (U, At = C \cup \{D\}, \{V_a\}, \{I_a\})$*,* $R \subseteq C$ *is an attribute reduct if and only if*

(1) $R = \arg\min_{R \subseteq C} \{\mathbf{COST}_R\}$*,*
(2) $\forall R' \subset R$*,* $\mathbf{COST}_{R'} > \mathbf{COST}_R$*.*

Compared to other definitions of attribute reduction, the main difference is that our definition is irrelevant to the positive region or non-negative region, and the objective of the reduction is to help users make better decisions which means the overall decision cost is minimum.

3.3 Generality of the Optimization Problem

From the definition of the decision cost, we know that the cost is composed of three parts: cost of positive rules, cost of boundary rules and cost of negative rules. In Equation(8), the importance of three types of costs is considered as same. In some actual applications, it may not be suitable, e.g. a user prefers positive rules and negative rules, which means he wants to make a direct decision. Another user may prefer positive rules and boundary rules. For this kind of situation, we propose a generality of the optimization problem, which will help users get a proper result.

$$\mathbf{COST} = \epsilon_P \cdot \sum_{p_i \geq \alpha} (1-p_i) \cdot \lambda_{PN} + \epsilon_B \cdot \sum_{\beta < p_j < \alpha} (p_j \cdot \lambda_{BP} + (1-p_j) \cdot \lambda_{BN}) + \epsilon_N \cdot \sum_{p_k \leq \beta} p_k \cdot \lambda_{NP}, \tag{14}$$

where ϵ_P, ϵ_B and ϵ_N denote the penalty of each kind of rules, respectively. If we set $\epsilon_P = \epsilon_N = 1$ and $\epsilon_B > 1$, then the resolution of the optimization problem is prefer to make less boundary rules. More different settings can be discussed in corresponding applications. We need more expert opinions or experiments to decide the exact values of the three penalty functions.

4 Conclusion

In this paper we present an optimization viewpoint on decision-theoretic rough set model. The decision cost is the basis of the model, and all rules are generated based on the minimum decision cost. From the view of that, we build an optimization problem in decision-theoretic rough set model with the objective of minimizing the decision cost. Through solving the optimization problem, we can learn cost functions and thresholds from data without any preliminary knowledge. This is an important attempt on learning cost functions automatically. We also define an attribute reduction based on the optimization problem. An attribute reduct is a minimal attribute set, and decisions generated from that will bring minimum costs. We can see that the optimization viewpoint could bring some new insights into decision-theoretic rough set model.

Acknowledgements. This work is supported by Natural Science Foundation of Jiangsu Province under Grant No. BK2009233.

References

1. Herbert, J.P., Yao, J.T.: Game-Theoretic Risk Analysis in Decision-Theoretic Rough Sets. In: Wang, G., Li, T., Grzymala-Busse, J.W., Miao, D., Skowron, A., Yao, Y. (eds.) RSKT 2008. LNCS (LNAI), vol. 5009, pp. 132–139. Springer, Heidelberg (2008)
2. Pawlak, Z., Wong, S.K.M., Ziarko, W.: Rough sets: probabilistic versus deterministic approach. International Journal of Man-machine Studies 29, 81–95 (1988)
3. Slezak, D., Ziarko, W.: The investigation of the Bayesian rough set model. International Journal of Approximate Reasoning 40, 81–91 (2005)
4. Yao, Y.Y., Wong, S.K.M.: A decision theoretic framework for approximating concepts. International Journal of Man-machine Studies 37(6), 793–809 (1992)
5. Yao, Y.Y.: Probabilistic approach to rough sets. Expert Systems 20, 287–297 (2003)
6. Yao, Y.Y.: Probabilistic rough set approximations. International Journal of Approximate Reasoning 49, 255–271 (2008)
7. Yao, Y.Y.: Three-way decisions with probabilitic rough sets. Information Sciences 180, 341–353 (2010)
8. Yao, Y., Zhou, B.: Naive Bayesian Rough Sets. In: Yu, J., Greco, S., Lingras, P., Wang, G., Skowron, A. (eds.) RSKT 2010. LNCS(LNAI), vol. 6401, pp. 719–726. Springer, Heidelberg (2010)
9. Yao, Y.Y.: The superiority of three-way decisions in probabilistic rough set models. Information Sciences 180(6), 1080–1096 (2011)
10. Zhao, Y., Wong, S.K.M., Yao, Y.Y.: A Note on Attribute Reduction in the Decision-Theoretic Rough Set Model. In: Chan, C.-C., Grzymala-Busse, J.W., Ziarko, W.P. (eds.) RSCTC 2008. LNCS (LNAI), vol. 5306, pp. 61–70. Springer, Heidelberg (2008)
11. Ziarko, W.: Variable precision rough set model. Journal of Computer and System Science 46, 39–59 (1993)