IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 13, 2020, accepted November 27, 2020, date of publication December 7, 2020, date of current version December 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042799

Decision-Theoretic Rough Set: A Fusion Strategy

TAO YIN[1], XIAOJUAN MAO[2], YING ZHANG[1], YITING MA[1], HENGRONG JU[1], AND WEIPING DING[1], (Senior Member, IEEE)

[1]School of Information Science and Technology, Nantong University, Nantong 226019, China
[2]Department of Respiratory Medicine, The Sixth People's Hospital of Nantong/Affiliated Nantong Hospital, Shanghai University, Nantong 216002, China

Corresponding author: Hengrong Ju (juhengrong@ntu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 62006128, Grant 61976120, and Grant 62076111; in part by the Jiangsu Innovation and Entrepreneurship Program; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20191445; in part by the Natural Science Foundation of Jiangsu Higher Education Institutions of China; and in part by the Qing Lan Project of Jiangsu Province and College Students' Innovative Entrepreneurial Training Plan Program under Grant 202010304023Z.

ABSTRACT Decision-theoretic rough set is a popular topic. However, such single-granulation rough set model is not able to handle complex information well, such as multi-source, multi-scale and high dimensions data. Therefore, the fusion of the ideas of Bayesian decision and multi-granulation may be an appealing issue. In this article, a novel rough set model based on multi-granularity decision theory is proposed. The discussed rough set model not only overcomes the shortcomings of optimistic and pessimistic rough sets, but also gains high approximation quality and low decision cost at the same time with a satisfactory threshold. In information granule reduction, heuristic and genetic algorithms are used to compute reducts based on three different criteria, respectively. The experimental results express that decision preservation based reduction may not suitable in such rough set models. Moreover, we also reveal that decision monotony and cost minimum based reductions are able to be popular research topics in rough set model of multi-granulation decision theory.

INDEX TERMS Decision cost, information granule reduction, multigranulation, optimization, rough set.

I. INTRODUCTION

Decision-theoretic Rough Set (DTRS), a special model of the traditional rough set [16], was put forward by Yao in the early 1990s. Different from classical rough set, DTRS introduces loss function and Bayesian decision procedure into rough set [13]. In DTRS, Yao used Bayesian theory to compute thresholds α and β, which was applied to depict error tolerance in probabilistic set by making the decision costs to a minimum. Based on DTRS, Yao studied numerous essential issues, which proposed an in-deep opinion of rough set. For instance, in [34], Yao offered a soil basis to DTRS's further studying. Then, Yao further presented the notion of three-way decisions and studied its superiority in practical applications [35]. In [36], Yao examined two fundamental semantics-related questions. Based on his seminal work, a lot of theoretical and practical achievements connected with DTRS have been obtained, see [2], [8]–[10], [23], [26], [28], [38], [40] for more details.

The associate editor coordinating the review of this manuscript and approving it for publication was Hao Ji.

In the perspective of the Granular Computing (GrC) [31], the set of data granules derived from one binary relation can be referred to as the granular structure [15], [27]. Since one binary relation can induce one granular structure in the universe, Yao's DTRS can be considered as a single-granulation rough set. Then it should be noted that Qian *et al.* argued that in numerous practical applications, single granular structure is not good enough for us to deal with complex data [14], [17], [25]. In order to solve it, Qian *et al.* [19], [20] put forward an innovative theory, which is named as the Multi-Granulation Rough Set (MGRS). In MGRS, the target is approximated by a multi-granulation structure. Presently, many of the positive results about MGRS have been achieved, e.g. [5], [11], [12], [14], [22], [24], [30].

With the rapid developments of decision-theoretic rough set and multi-granulation rough set, how to fuse Bayesian decision and multi-granulation ideas has become a valuable problem. In numerous practical applications, it is necessary to take different granulations into consideration when one makes an important decision, such as multi-scale data analysis, multi-source data analysis, face recognition. In this article, we introduce following several instances to emphasize the incentive which inspire us to consider the Bayesian decision and multi-granulation simultaneously.

1) In the process of knowledge discovery, many types of data sets are existed, including categorical information, wording information, picture data, audio information, and the like [12], [29]. For instance, identifying the violence video from the Internet is one of the security personnel's major duties. A conventional video should include frame information. During the course of such decision making, the security person should make judgements from those three aspects, respectively.
2) In many real-life applications, under a specific feature, an object can be denoted as many values with respect to different scales [31]. For example, students' test scores can be recorded as positive numbers from 0 to 100, and sometimes, for simplicity, the scores might be divided into two values, "Pass" and "Fail". In this case, the decision mechanisms will be different in different scales.
3) When encountering information with high dimensions, many attributes lead to a challenge for decision making. For example, we note that, the data sets with high dimensions, high noise or small sample are ubiquitous in biomedicine field, especially some applications based on Omics data, DNA translation initiation detection, activity prognosis using medicine molecules information and so on [1]. In this case, if we granulate the data using all attributes, it will resulted in time-consuming.

From such point of view, Qian *et al.* [21] brought in the multi-granulation idea into decision-theoretic rough set model at first attempt, i.e., Multi-Granulation Decision-Theoretic Rough Set (MG-DTRS). As we all know, the defined MG-DTRS includes optimistic and pessimistic forms, which are only two simple methods in information fusion strategies [3] and lack flexibility while simulating the brain's intelligence in decision making. There is an example can be presented to illustrate the limitations of optimistic and pessimistic versions. The political subdivision of China has many provinces. In order to show the economical development status of China, the optimistic experts will focus on the regions of Shanghai Provence or the others in Eastern China, based on the GDP of these regions, the China can be regarded as a developed country. The pessimistic experts will focus on the regions, such as Gansu Province or the others in Northwest China, based on the GDP of these regions, the China can be regarded as an impoverished nation. However, such two conclusions both do not conform to China's actual conditions. Therefore, it is important to present a new model to extend the optimistic and pessimistic versions. This is what will be explored in this article.

The rest of the paper is structured as below. Rudimentary knowledge, like multi-granulation decision-theoretic rough set, Yao's DTRS will be presented in Section II. In Section III, we will propose a new multi-granulation decision-theoretic rough set, which is called δ-Multi-granulation Decision-theoretic Rough Set. In Section IV, multi-granulation data reduction will be considered by three criteria. Section V discusses the efficiency of proposed algorithms. The paper ends up with conclusions in Section VI.

II. PRELIMINARY KNOWLEDGE

A. DECISION-THEORETIC ROUGH SET

In order to introduce the cost mechanism into rough set model, Yao *et al.* proposed a neoteric rough set theory. The new rough set is called as decision-theoretic rough set. In DTRS, the decision process is described by two states and three actions. With respect to a target X in the universe, the two states is expressed by $\Omega = \{X, X^c\}$. The set of actions regarding the state X is given by $T = \{e_P, e_B, e_N\}$, where e_P, e_B, e_N denote the actions in deciding an object into positive region, boundary region and negative region, respectively. The loss function with respect to the costs of different actions can be presented as following matrix.

	$X(P)$	$X^c(N)$
e_P	λ_{PP}	λ_{PN}
e_B	λ_{BP}	λ_{BN}
e_N	λ_{NP}	λ_{NN}

According to the loss function, the expected losses for distinctive actions are able to be represented as:

$$R_P = R(e_P|[x]_A) = \lambda_{PP} \cdot P(X|[x]_A) + \lambda_{PN} \cdot P(X^c|[x]_A);$$
$$R_B = R(e_B|[x]_A) = \lambda_{BP} \cdot P(X|[x]_A) + \lambda_{BN} \cdot P(X^c|[x]_A);$$
$$R_N = R(e_N|[x]_A) = \lambda_{NP} \cdot P(X|[x]_A) + \lambda_{NN} \cdot P(X^c|[x]_A).$$

Furthermore, for the loss settings in loss function, exceptional circumstances are able to be proposed as below:

$$\lambda_{PP} \le \lambda_{BP} < \lambda_{NP}, \quad \lambda_{NN} \le \lambda_{BN} < \lambda_{PN}$$
$$(\lambda_{PN} - \lambda_{BN}) \cdot (\lambda_{NP} - \lambda_{BP}) > (\lambda_{BP} - \lambda_{PP}) \cdot (\lambda_{BN} - \lambda_{NN})$$

Based on such circumstance, the rule of determining the minimal danger is able to be expressed as:

(P) If $P(X|[x]_A) \ge \alpha$, then decides positive region;
(B) If $\beta < P(X|[x]_A) < \alpha$, then decides boundary region;
(N) If $P(X|[x]_A) \le \beta$, then decides negative region;
where

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

with $0 \le \beta < \alpha \le 1$.

By operating three decision rules, the lower and upper approximations of the DTRS are able to be gained, thus:

$$\underline{A}_{DT}(X) = \{x \in U : P(X|[x]_A) \ge \alpha\}; \tag{1}$$
$$\overline{A}_{DT}(X) = \{x \in U : P(X|[x]_A) > \beta\}. \tag{2}$$

B. MULTI-GRANULATION DECISION-THEORETIC ROUGH SET

Formally, a multi-granulation decision system (MDS) [14] is: $S =< U, \mathcal{AT} = \{A_1, \ldots, A_m\}, D, \{V_a : a \in A_1 \cup A_2 \cup \ldots \cup A_m \cup D\} >$, in which

- Universe U is a finite set of objects;
- A_k $(k = 1, 2, \ldots, m)$ is a set of the condition data, m is a natural number;
- D is the set of the decision data;
- $\forall a \in A_1 \cup A_2 \cup \ldots \cup A_m \cup D$, V_a is the set of values for a.

Different from traditional decision system, $\forall A_k \in \mathcal{AT}$, one can obtain partition, i.e., information granulation and then $\mathcal{AT}$ actually induces a multi-granulation information.

In multi-granulation decision system, Qian *et al.* have developed a novel rough set model through fusing the idea of multi-granulation and DTRS theory, i.e., MG-DTRS. In this article, we use the assumption that every information granule corresponding to an independent loss function. The loss function of the $k-$th data granulation is also given by the matrix as follows:

	$X(P)$	$X^c(N)$
e_P	λ^k_{PP}	λ^k_{PN}
e_B	λ^k_{BP}	λ^k_{BN}
e_N	λ^k_{NP}	λ^k_{NN}

Based on the special kind of loss function, the decided parameters for $k-$th data granulation are expressed by α^k, β^k, $k \in \{1, 2, \ldots, m\}$.

Definition 1: Let S be a MDS, $\forall X \subseteq U$, the optimistic multigranulation decision-theoretic lower / upper approximations are denoted by $\underline{\mathcal{AT}}_{ODT}(X)$ and $\overline{\mathcal{AT}}_{ODT}(X)$, respectively,

$$\underline{\mathcal{AT}}_{ODT}(X) = \{x \in U : P(X|[x]_{A_k}) \geq \alpha^k, \exists A_k \in \mathcal{AT}\}; \quad (3)$$
$$\overline{\mathcal{AT}}_{ODT}(X) = \{x \in U : P(X|[x]_{A_k}) > \beta^k, \forall A_k \in \mathcal{AT}\}. \quad (4)$$

Based on the lower approximation and upper approximation, the optimistic boundary region of X is presented as:

$$BN_{ODT}(X) = \overline{\mathcal{AT}}_{ODT}(X) - \underline{\mathcal{AT}}_{ODT}(X). \quad (5)$$

Definition 2: Let S be a MDS, $\forall X \subseteq U$, the pessimistic multi-granulation decision-theoretic lower/upper approximations are expressed by $\underline{\mathcal{AT}}_{PDT}(X)$ and $\overline{\mathcal{AT}}_{PDT}(X)$, respectively,

$$\underline{\mathcal{AT}}_{PDT}(X) = \{x \in U : P(X|[x]_{A_k}) \geq \alpha^k, \forall A_k \in \mathcal{AT}\}; \quad (6)$$
$$\overline{\mathcal{AT}}_{PDT}(X) = \{x \in U : P(X|[x]_{A_k}) > \beta^k, \exists A_k \in \mathcal{AT}\}. \quad (7)$$

Based on the lower approximation and upper approximation, the pessimistic boundary region of X is presented as:

$$BN_{PDT}(X) = \overline{\mathcal{AT}}_{PDT}(X) - \underline{\mathcal{AT}}_{PDT}(X). \quad (8)$$

III. δ-MULTI-GRANULATION DECISION-THEORETIC ROUGH SET (δ-MGDTRS)

A. DEFINITION OF δ-MGDTRS

Definition 3: Let S be a MDS, then $\forall x \in U$ and $\forall X \subseteq U$, for each $k \in \{1, 2, \ldots, m\}$, the characteristic functions are defined as

$$f^k_X(x) = \begin{cases} 1 & : P(X|[x]_{A_k}) \geq \alpha^k \\ 0 & : \text{otherwise} \end{cases}$$
$$\psi^k_X(x) = \begin{cases} 1 & : P(X|[x]_{A_k}) > \beta^k \\ 0 & : \text{otherwise} \end{cases}$$

$f^k_X(x)$ and $\psi^k_X(x)$ judge whether the $k-$th information granulation meets the condition probability constraints $\geq \alpha^k$ and $> \beta^k$, respectively.

Definition 4: Let S be a MDS, $f^k_X(x)$ and $\psi^k_X(x)$ are the supporting characteristic functions of each x in U, then the lower approximation and upper approximation are expressed by $\underline{\mathcal{AT}}_{\delta DT}(X)$ and $\overline{\mathcal{AT}}_{\delta DT}(X)$, respectively,

$$\underline{\mathcal{AT}}_{\delta DT}(X) = \{x \in U : \frac{\sum_{k=1}^m f^k_X(x)}{m} \geq \delta\}; \quad (9)$$
$$\overline{\mathcal{AT}}_{\delta DT}(X) = \{x \in U : \frac{\sum_{k=1}^m \psi^k_X(x)}{m} > 1 - \delta\}; \quad (10)$$

where $\delta \in (0, 1]$.

The pair $[\underline{\mathcal{AT}}_{\delta DT}(X), \overline{\mathcal{AT}}_{\delta DT}(X)]$ is considered as a δ-MGDTRS of X. According to Definition 4, one is able to notice that objects which are located in lower approximation if and only if the number of the conditional probability constraints ($\geq \alpha^k$) is no less than $m \times \delta$. Similarity, objects which are located in upper approximation if and only if the number of the conditional probability constraints ($> \beta^k$) is more than $m \times (1 - \delta)$.

By $\underline{\mathcal{AT}}_{\delta DT}(X)$ and $\overline{\mathcal{AT}}_{\delta DT}(X)$, the three regions in δ-multi-granulation decision-theoretic environment are denoted as:

$$POS_\delta(\mathcal{AT}, X) = \underline{\mathcal{AT}}_{\delta DT}(X); \quad (11)$$
$$BND_\delta(\mathcal{AT}, X) = \overline{\mathcal{AT}}_{\delta DT}(X) - \underline{\mathcal{AT}}_{\delta DT}(X); \quad (12)$$
$$NEG_\delta(\mathcal{AT}, X) = U - POS_\delta(\mathcal{AT}, X) \cup BND_\delta(\mathcal{AT}, X) = U - \overline{\mathcal{AT}}_{\delta DT}(X). \quad (13)$$

Definition 5: Let S be a MDS, $\delta \in (0, 1]$, the approximation quality of D based on $\delta-$MGDTRS is charactersized as below:

$$\gamma(\delta, D) = \frac{|\bigcup_{j=1}^n \underline{\mathcal{AT}}_{\delta DT}(X_j)|}{|U|}. \quad (14)$$

B. RELATED PROPERTIES

Proposition 1: Let S be a MDS, $\delta \in (0, 1]$, $\forall X \subseteq U$, we have

$$\underline{\mathcal{AT}}_{PDT}(X) \subseteq \underline{\mathcal{AT}}_{\delta DT}(X) \subseteq \underline{\mathcal{AT}}_{ODT}(X); \quad (15)$$
$$\overline{\mathcal{AT}}_{ODT}(X) \subseteq \overline{\mathcal{AT}}_{\delta DT}(X) \subseteq \overline{\mathcal{AT}}_{PDT}(X). \quad (16)$$

Proposition 1 indicates that $\delta-$multi-granulation decision-theoretic lower approximation is equal to or greater than pessimistic multi-granulation decision-theoretic lower approximation, it is also not lager than optimistic multi-granulation lower approximation of decision theory, $\delta-$multi-granulation upper approximation of decision theory is not smaller than optimistic multi-granulation decision-theoretic upper approximation, it is also not more than pessimistic multi-granulation decision-theoretic upper approximation.

Proposition 2: Let S be a MDS, if $0 < \delta_1 < \delta_2 \leq 1$, then $\forall X \subseteq U$, we have

$$\underline{\mathcal{AT}}_{\delta_2 DT}(X) \subseteq \underline{\mathcal{AT}}_{\delta_1\ DT}(X); \tag{17}$$

$$\overline{\mathcal{AT}}_{\delta_1 DT}(X) \subseteq \overline{\mathcal{AT}}_{\delta_2 DT}(X). \tag{18}$$

Proposition 2 tells us that with the expanding of the value of δ, $\delta-$multi-granulation decision-theoretic lower approximation is decreasing while $\delta-$multi-granulation decision-theoretic upper approximation is increasing.

C. DECISION COST OF MULTI-GRANULATION DECISION SYSTEM

Let $U/IND(D) = \{X_1, X_2, \ldots, X_n\}$ become a division of the universe induced by D. In multi-granulation decision environment, $\forall X_j \in U/IND(D), j = \{1, 2, \ldots, n\}$, similar to the Yao's DTRS, one is able to gain determination principle tie-broke:

(δ-P) $\forall x \in U$, if $\frac{\sum_{k=1}^{m} f^k_{X_j}(x)}{m} \geq \delta$, $\delta \in (0, 1]$, then decides $x \in POS_\delta(\mathcal{AT}, X_j)$;

(δ-N) $\forall x \in U$, if $\frac{\sum_{k=1}^{m} \psi^k_{X_j}(x)}{m} \leq 1 - \delta$, $\delta \in (0, 1]$, then decides $x \in NEG_\delta(\mathcal{AT}, X_j)$;

(δ-B) Otherwise, then decides $x \in BND_\delta(\mathcal{AT}, X_j)$.

By using DTRS model, Yao discussed many significant and essential measures so as to assess the performances of classification rules. One of the important achievements of Yao's DTRS is introducing decision cost into rough set model. Therefore, decision cost has been a most popular measure to evaluate the decision rules' performance. As a generalization of Yao's DTRS, decision cost plays a key role in our δ-MGDTR as well.

Let S be a MDS, $\delta \in (0, 1]$, $\forall X_j \in U/IND(D)$, the Bayesian expected decision cost can be expressed as below:

$$\mathrm{COST}_{POS} = \sum_{X_j \in U/IND(D)} \sum_{x \in POS_\delta(X_j)} \sum_{k=1}^{m} \Big(\lambda^k_{PP} \cdot P(X_j|[x]_{A_k}) + \lambda^k_{PN} \cdot P(X_j^c|[x]_{A_k})\Big);$$

$$\mathrm{COST}_{NEG} = \sum_{X_j \in U/IND(D)} \sum_{x \in NEG_\delta(X_j)} \sum_{k=1}^{m} \Big(\lambda^k_{NP} \cdot P(X_j|[x]_{A_k}) + \lambda^k_{NN} \cdot P(X_j^c|[x]_{A_k})\Big);$$

$$\mathrm{COST}_{BND} = \sum_{X_j \in U/IND(D)} \sum_{x \in BND_\delta(X_j)} \sum_{k=1}^{m} \Big(\lambda^k_{BP} \cdot P(X_j|[x]_{A_k}) + \lambda^k_{BN} \cdot P(X_j^c|[x]_{A_k})\Big).$$

The whole decision cost is:

$$COST(\mathcal{AT}, \delta) = \mathrm{COST}_{POS} + \mathrm{COST}_{NEG} + \mathrm{COST}_{BND} \tag{19}$$

D. A NAIVE APPROACH TO LEARNING THRESHOLD

In our paper, a threshold δ is introduced to extend Qian *et al.*'s MG-DTRS. With the value of δ changing, the new MG-DTRS can be converted to Qian *et al.*'s MG-DTRS. Therefore, the way to select a fitness threshold to construct a satisfactory MG-DTRS model is an important issue. Firstly, one must figure out what is the definition of satisfaction. Different opinions can be obtained based on different considerations. From the viewpoint of classification, the satisfactory result is that one can improve the classification accuracy, whereas, from the viewpoint of risk decision, the satisfactory result is that one can reduce the decision cost. As far as MG-DTRS is concerned, the approximation quality and decision cost are two important aspects which should be seriously considered. On the one hand, the concept of approximation quality not only can be regarded as the probability of objects which can be accurately divided into the set of lower approximation, but also is one quantitative descriptions of positive decision rule. On the other hand, the effectiveness of risk decision procedure is the reflection of the decision cost. From this point of view, a satisfactory MG-DTRS model can be defined with high approximation quality and low decision cost.

To fill such gap, a fusion function for fusing the Approximation Quality changing Ratio (*AQR*) with Decision Cost changing Ratio (*DCR*) can be considered. *AQR* is the proportion of the approximate quality obtained by the current threshold to the largest approximate quality, which reflects the influence of the current threshold on the approximate quality. Similar to *AQR*, *DCR* is used to measure the impact of current threshold on decision cost. The fusion function (*ADR*) can be defined as:

$$\begin{aligned} ADR(\delta_i) &= w_1 \cdot AQR - w_2 \cdot DCR \\ &= w_1 \cdot \frac{\gamma(\delta_i, D)}{\max \gamma(\delta, D)} - w_2 \cdot \frac{COST(\mathcal{AT}, \delta_i)}{\max COST(\mathcal{AT}, \delta)} \end{aligned} \tag{20}$$

in which δ_i is a arbitrary δ in (0,1], the weights $w_1 > 0$, $w_2 > 0$. Setting these two weights need some technical support. In this article, from a simple point of view, we can make a hypothesis that AQR and DCR have equal statuses and set as $w_1 = 1, w_1 = 1$.

In (20), one can observe that, for a certain data set, with different values of δ, we can generate different MG-DTRS models, each MG-DTRS model owns certain approximation quality and decision cost. Maximal values of approximation quality and decision cost can be obtained from the

Algorithm 1 A Naive Threshold Learning Algorithm to Learning Satisfactory MG-DTRS Model (NTLA)

Input: MDS S, step l;
Output: A satisfactory threshold δ_s.
Step 1: Compute $U/IND(D)$;
Step 2: $\underline{\mathcal{AT}}_{\delta DT}(X_j) = \emptyset$, $\overline{\mathcal{AT}}_{\delta DT}(X_j) = \emptyset$, $X_j \in U/IND(D)$; $\delta_s = \delta = 0$.
Step 3: Generate the loss function and compute $\alpha_k, \beta_k, k = \{1, 2, \ldots, m\}$;
Step 4: **For** each $x \in U$
  For $i = 1$ to l
    $\delta = \delta + 1/l$;
    For $j = 1$ to n
      Compute $f^k_{X_j}(x)$, $\psi^k_{X_j}(x)$, according to Definition 3, $k = 1, 2, \ldots, m$;
      If $\sum_{k=1}^{m} f^k_{X_j}(x)/m \geq \delta$
        $\underline{\mathcal{AT}}_{\delta DT}(X_j) = \underline{\mathcal{AT}}_{\delta DT}(X_j) \cup \{x\}$;
      End If
      If $\sum_{k=1}^{m} \psi^k_{X_j}(x)/m > 1 - \delta$
        $\overline{\mathcal{AT}}_{\delta DT}(X_j) = \overline{\mathcal{AT}}_{\delta DT}(X_j) \cup \{x\}$;
      End If
    End For
    Compute $\gamma(\delta, D)$ and COST$(\mathcal{AT}, \delta)$;
  End For
  End For
Step 5: **For** $i = 1$ to l
  Find the maximum values of $\gamma(\delta_i, D)$ and COST$(\mathcal{AT}, \delta_i)$;
  Compute $ADR(\delta_i)$;
  End For
Step 6: Return the maximum value of $ADR(\delta)$ and the satisfactory δ_s;

approximation quality set and decision cost set, respectively. Therefore, with different values of δ, we can obtain different ADR values. Furthermore, with the increasing of AQR and decreasing of DCR, the evaluation criterion ADR will be increasing. In light of the above discussions, the satisfactory MG-DTRS model or satisfactory threshold is one which holds a maximum value of ADR. Obviously, the obtained δ is not a optimal value. However, considering the theory of Ockahm's razor, we endeavor to build a simple model to balance the requirements of these two considerations and simplicity.

According to these discussions, we can design a naive threshold learning algorithm as follows:

IV. MULTI-GRANULATION INFORMATION REDUCTION

A. DECISION BASED INFORMATION GRANULE REDUCTION

Definition 6: Let S be a MDS, $\delta \in (0, 1]$, for $\mathcal{B} \subseteq \mathcal{AT}$, $\mathcal{B}$ is called a Decision Preservation (DP) reduction when and only when

1) $\mathcal{B}$ is a decision preservation consistent set in S, i.e. $\underline{\mathcal{AT}}_{\delta DT}(X_j) = \underline{\mathcal{B}}_{\delta DT}(X_j)$, $X_j \in U/IND(D)$;
2) For each $\mathcal{B}' \subset \mathcal{B}$, $\underline{\mathcal{B}'}_{\delta DT}(X_j) \neq \underline{\mathcal{B}}_{\delta DT}(X_j)$, $X_j \in U/IND(D)$.

In Definition 6, it can be observed that the decision preservation(DP) reduct in S is the smallest subset of $\mathcal{AT}$, it keeps the $\delta-$multi-granulation decision-theoretic lower approximations of the whole decision classes stable as well. In addition, note that positive determination principles are supported by targets in positive region (lower approximation), then from Definition 6, decision preservation (DP) reduct will not change the numbers of positive decision principles. In other words, the decision rules derived from the decision preservation consistent set $\mathcal{B}$ are compatible with the ones derived from $\mathcal{AT}$.

Definition 7: Let S be a MDS, $\delta \in (0, 1]$, for each $\mathcal{B} \subseteq \mathcal{AT}$, $\mathcal{B}$ is called a Decision Monotonicity (DM) reduct when and only when

1) $\mathcal{B}$ is a decision monotonicity consistent set in S, i.e. $\underline{\mathcal{AT}}_{\delta DT}(X_j) \subseteq \underline{\mathcal{B}}_{\delta DT}(X_j)$, $X_j \in U/IND(D)$;
2) For each $\mathcal{B}' \subset \mathcal{B}$, $\underline{\mathcal{B}'}_{\delta DT}(X_j) \not\supseteq \underline{\mathcal{B}}_{\delta DT}(X_j)$, $X_j \in U/IND(D)$.

According to the decision monotonicity (DM) reduct, condition (1) can be seen as a sufficient situation while condition (2) can be regarded as a necessary situation. The sufficient condition illustrates that not only does decision monotonicity based reduct keep the original positive rules stable, but increase the number of positive rules as more as possible.

Proposition 3: Let S be a MDS, for each $\mathcal{B} \subseteq \mathcal{AT}$, if $\mathcal{B}$ is a decision preservation consistent set in S, then $\mathcal{B} \subseteq \mathcal{AT}$ is a decision monotonicity consistent set in S.

B. COST BASED INFORMATION GRANULE REDUCTION

Risk cost is a typical and important feature in the multi-granulation decision framework. In above Subsection, one have discussed the cost issue of our MG-DTRS. However, in the reduction process, it must be noticed that there are more than one reduct in real world applications. From the perspective of minimum-risk, one may want to minimize cost for a maximized reduct. The cost based information granule reduction can be seen as below.

Definition 8: Let S be a MDS, $\delta \in (0, 1]$, for each $\mathcal{B} \subseteq \mathcal{AT}$, $\mathcal{B}$ is referred to as a Cost Minimum (CM) reduct when and only when

1) $\mathcal{B}$ is a cost minimum consistent set in S, i.e. COST$(\mathcal{B}, \delta) \leq$ COST$(\mathcal{AT}, \delta)$;
2) For each $\mathcal{B}' \subset \mathcal{B}$, COST$(\mathcal{B}', \delta) >$ COST$(\mathcal{B}, \delta)$.

In definition 8, condition (1) and condition (2) are able to be separately considered as sufficient condition and necessary condition. From the perspective of optimization, the minimum decision cost is able to be expressed [4]:

$$\min COST(\mathcal{B}, \delta). \tag{21}$$

C. HEURISTIC AND OPTIMIZATION INFORMATION GRANULE REDUCTION ALGORITHMS

1) HEURISTIC INFORMATION GRANULE REDUCTION ALGORITHMS

We should take three important problems into consideration about heuristic algorithm based information granule reduction, [18], i.e., importance measures of information granulation, search strategy and termination criterion. In rough set theory, the definition of significance of information granule plays a crucial role for the successful information granule reduction. As a result, in the forward greedy search approach, heuristic functions use two significant information granule's measures to support efficient information granule reduction, which are also separately named as inner importance measure and outer importance measure. As is shown, three types of significance will be presented to measure the importance of information granule.

Definition 9: Let S be a MDS, $\delta \in (0, 1]$, $\mathcal{B} \subseteq \mathcal{AT}$, $\forall A_k \in \mathcal{B}$, the decision preservation based significance of A_k in $\mathcal{B}$ can be defined as follows:

$$Sig_1^{inner}(A_k, \mathcal{B}, D) = \frac{\sum_{j=1}^{n}\{|\underline{\mathcal{B}}_{\delta DT}(X_j) \oplus \underline{\mathcal{B}-\{A_k\}}_{\delta DT}(X_j)|\}}{|U|}; \tag{22}$$

where $X \oplus Y$ expresses the symmetric difference of X and Y.

Definition 10: Let S be a MDS, $\delta \in (0, 1]$, $\mathcal{B} \subseteq \mathcal{AT}$, $\forall A_k \in \mathcal{AT} - \mathcal{B}$, the decision preservation based significance of A_k which is out of $\mathcal{B}$ can be defined as shown:

$$Sig_1^{outer}(A_k, \mathcal{B}, D) = \frac{\sum_{j=1}^{n}\{|\underline{\mathcal{B}}_{\delta DT}(X_j) \oplus \underline{\mathcal{B} \cup \{A_k\}}_{\delta DT}(X_j)|\}}{|U|}. \tag{23}$$

$Sig_1^{inner}(A_k, \mathcal{B}, D)$ is the decision preservation based significance of A_k in $\mathcal{B}$ connected with decision D. Obviously, $Sig_1^{inner}(A_k, \mathcal{B}, D) \geq 0$ and $Sig_1^{outer}(A_k, \mathcal{B}, D) \geq 0$ hold. At the same time, the greater the value is, the more important information granule A_k is. Furthermore, if $Sig_1^{inner}(A_k, \mathcal{B}, D) > 0$, then A_k is a core information granule of S [18].

Remark 1: It should be noticed that our MG-DTRS is not monotonic variation with monotonic variation of granular structures. Therefore, we used symmetric difference to describe the variation of approximations. This is different from the previous measure of significance [6].

For decision monotonicity based information granule reduction, Proposition 3 illustrates a relationship between decision preservation consistent set and decision monotonicity consistent set. For example, for each $\mathcal{B} \subseteq \mathcal{AT}$, if $\mathcal{B}$ is a decision preservation consistent set in S, then $\mathcal{B} \subseteq \mathcal{AT}$ is a decision monotonicity consistent set in S. The only difference is that the decision monotonicity consistent set needs to increase the decision region monotonously on the basis of decision preservation criterion. In other words, from the viewpoint of heuristic algorithm, it may have some difference in the outer significance measure applied to a forward information granule reduction. Meanwhile, the inner important measure holds the same function. It means that the core information granule does not change. In the following, a new measure will be put forward.

Definition 11: Let S be a MDS, $\delta \in (0, 1]$, $\mathcal{B} \subseteq \mathcal{AT}$, $\forall A_k \in \mathcal{AT} - \mathcal{B}$, the decision monotonicity based significance of A_k which is out of $\mathcal{B}$ can be defined as follows:

$$Sig_2^{outer}(A_k, \mathcal{B}, D) = \frac{\sum_{j=1}^{n}\left(\underline{\mathcal{B}}_{\delta DT}(X_j) \odot \underline{\mathcal{B} \cup \{A_k\}}_{\delta DT}(X_j)\right)}{|U|}. \tag{24}$$

where $X \odot Y$ is an operation such that:

$$X \odot Y = \begin{cases} |Y - X| & X \subseteq Y \\ -\infty & otherwise \end{cases}$$

By means of the operation "$\odot$", the decision monotonicity based on significance can be obtained. The value of $Sig_2^{outer}(A_k, \mathcal{B}, D)$ may be no less than zero or minus infinity. $Sig_2^{outer}(A_k, \mathcal{B}, D)$ indicates the decision monotonicity based on significance of A_k, which is introduced into the set $\mathcal{B}$.

For decision cost based information granule reduction, the main idea is to decrease the decision cost to the lowest. Following the mechanism of heuristic algorithm, we make a new definition of the inner importance measure and outer importance measure as shown.

Definition 12: Let S be a MDS, $\delta \in (0, 1]$, $\mathcal{B} \subseteq \mathcal{AT}$, $\forall A_k \in \mathcal{B}$, the cost minimum based significance of A_k in $\mathcal{B}$ is able to be expressed as shown:

$$Sig_3^{inner}(A_k, \mathcal{B}, D) = \frac{\mathrm{COST}(\mathcal{B} - \{A_k\}, \delta) - COST(\mathcal{B}, \delta)}{\mathrm{COST}(\mathcal{B}, \delta)}; \tag{25}$$

Definition 13: Let S be a MDS, $\delta \in (0, 1]$, $\mathcal{B} \subseteq \mathcal{AT}$, $\forall A_k \in \mathcal{AT} - \mathcal{B}$, the cost minimum based significance of A_k in $\mathcal{B}$ can be defined as shown:

$$Sig_3^{outer}(A_k, \mathcal{B}, D) = \frac{\mathrm{COST}(\mathcal{B}, \delta) - \mathrm{COST}(\mathcal{B} \cup \{A_k\}, \delta)}{\mathrm{COST}(\mathcal{B}, \delta)}. \tag{26}$$

$Sig_3^{inner}(A_k, \mathcal{B}, D)$ and $Sig_3^{outer}(A_k, \mathcal{B}, D)$ are the decision cost based significance measures. The value of $Sig_3^{inner}(A_k, \mathcal{B}, D)$ shows that the decision cost change when we delete an information granule from the current set. $Sig_3^{inner}(A_k, \mathcal{B}, D) > 0$ means that the decision cost has been increased, which indicates that A_k should not be deleted. From the perspective of decision cost, A_k can also be regarded as a core information granule of S. $Sig_3^{inner}(A_k, \mathcal{B}, D) = 0$ illustrates that the decision cost remains unchanged when A_k is deleted from $\mathcal{B}$. As a result, A_k is an irrelevant information granule and should be removed. From the discussion above, one can observe the greater the value is, the more significant information granule $A_k$is. Accordingly, similar explanation of $Sig_3^{outer}(A_k, \mathcal{B}, D)$ can be obtained, it also shows that the greater the value of $Sig_3^{outer}(A_k, \mathcal{B}, D)$ is, the more significant information granule A_k is.

Algorithm 2 Heuristic Algorithm for Information Granule Reduct (HAIGR)

Input: MDS S, threshold δ;
Output: A reduct red.
Step 1: $\forall X_j \in U/IND(D)$, compute $\underline{\mathcal{AT}}_{\delta DT}(X_j)$;
Step 2: $\mathcal{B} \leftarrow \emptyset$;
Step 3: Compute the significance for each $A_k \in \mathcal{AT}$ with $Sig_{\bullet}^{inner}(A_k, \mathcal{AT}, D)$;
Step 4: $\mathcal{B} \leftarrow A_j$ where $Sig_{\bullet}^{inner}(A_j, \mathcal{AT}, D) = \max\{Sig_{\bullet}^{inner}(A_k, \mathcal{AT}, D) : A_k \in \mathcal{AT}\}$;
Step 5:
  Do
    $\forall A_k \in \mathcal{AT} - \mathcal{B}$, compute $Sig_{\bullet}^{outer}(A_k, \mathcal{AT}, D)$;
    If $Sig_{\bullet}^{outer}(A_j, \mathcal{AT}, D) = \max\{Sig_{\bullet}^{outer}(A_k, \mathcal{AT}, D) : \forall A_k \in \mathcal{AT} - \mathcal{B}\}$
    $\mathcal{B} = \mathcal{B} \cup \{A_j\}$;
    End
    Until $\mathcal{B}$ is a decision preservation (decision monotonicity, cost minimum) consistent set in S;
Step 6: $\forall A_k \in \mathcal{B}$
  If $\mathcal{B} - A_k$ is decision preservation (decision monotonicity, cost minimum) consistent set
  $\mathcal{B} = \mathcal{B} - \{A_k\}$;
  End
Step 7: $red = \mathcal{B}$;

As far as search strategy is concerned, in each information granule reduction approach, the information granule with the maximal inner importance will be selected preferentially, then in each loop, the information granule with maximal outer significance will be brought into the information granule subset until it satisfies the stopping criterion. Formally, a heuristic information granule reduct algorithm is able to be expressed as shown.

In Algorithm 2, if $\bullet = 1$ and the stopping criterion is selected as ''decision preservation'', then the HAIGR is used to compute the decision preservation based reduct, which can be expressed as DP-HAIGR. Similarity, if $\bullet = 2$ and the stopping criterion is selected as ''decision monotonicity'', then the HAIGR is used to compute the decision preservation based reduct, which can be expressed as DM-HAIGR; if $\bullet = 3$ and the stopping criterion is selected as ''cost minimum'', then the HAIGR is used to compute the decision preservation grounded reduct, which can be expressed as CM-HAIGR.

From the above points, decision monotonicity based reduct and cost minimum based reduct are optimization problems. We use genetic algorithm reduction to achieve the purpose of optimization. Here, we introduce the fitness function for decision monotonicity based reduct and cost minimum based reduct.

Definition 14: Let S be a MDS, $\delta \in (0, 1]$, $\forall \mathcal{B} \subseteq \mathcal{AT}$, based on the determination monotonicity and cost minimal principle, the fitness function is able to expressed as below.

$$f^{DM} = \frac{\sum_{j=1}^{n} \left(\underline{\mathcal{AT}}_{\delta DT}(X_j) \odot \underline{\mathcal{B}}_{\delta DT}(X_j)\right)}{|U|}; \tag{27}$$

TABLE 1. Data sets description.

ID	Data Sets	Samples	Features	Classes
1	Abalone	4177	8	2
2	Adult4	4781	14	2
3	Dermatology	366	34	6
4	Molecular Biology	1242	60	2
5	Page Blocks	5473	10	5
6	Steel Plates Faults	1941	34	2
7	Wdbc	596	30	4
8	Zoo	101	17	7

$$f^{CM} = \frac{COST(\mathcal{B}, \delta)}{COST(\mathcal{AT}, \delta)}. \tag{28}$$

In terms of decision-monotonicity criterion, it is easy to find that $f^{DM} \in (-\infty, 1)$. $f^{DM} > 0$ indicates that the subset $\mathcal{B}$ obtains more positive decision rules than original information granule set $\mathcal{AT}$. $f^{DM} = 0$ shows that the positive determination principles induced by subset $\mathcal{B}$ are equal to the original; $f^{DM} < 0$ indicates the monotonicity has been lost between the two sets. What's more, in terms of cost minimum criterion, one is able to observe that $f^{CM} \in [0, 1]$, and the major purpose of cost minimal principle is to reduce the value of f^{CM}.

Genetic algorithm is a kind of evolutionary algorithm, and the pseudo-code can be described in [7]. In the algorithm, if the fitness function is chosen as "f^{DM}", then the genetic algorithm is applied to discover the best reduct based on decision monotonicity, namely DM-GAIGR. Similarity, if fitness function is selected as "f^{CM}", then genetic algorithm is applied to discover an optimal decision cost based reduct, namely CM-GAIGR.

V. EXPERIMENTAL ANALYSIS

In the part, effectiveness of discussed algorithms is shown by some experimentation. The data sets appeared are outlined in Table 1. Each attribute in a information set is applied to establish a part, i.e., information granule.

A. MG-DTRS MODELS' COMPARISONS

Firstly, in this subsection, the contrast among OMG-DTRS, PMG-DTRS and MGDTRS will be compared. Table 2 indicates approximation properties (i.e., AQ) of OMG-DTRS, PMG-DTRS and δ-MGDTRS in such eight data sets, respectively. Since δ-MGDTRS proposed in Definition 4 is limited by threshold δ, in order to defeat the shortcomings of empiricism, some different values of δ are selected. By Table 2, we can acquire two remarks.

1) Contrasted with these three MG-DTRS models, OMG- DTRS is able to acquire the max value of AQ and PMG-DTRS is able to acquire min value. The values of AQ based on δ-MGDTRS are between these of OMG-DTRS and PMG-DTRS.
2) In terms of variation of δ, it's easy to note that, with an increase in δ, the value of AQ based on δ-MGDTRS is decreasing. Such an experimental results demonstrate the theoretical ones shown in Proposition2.

TABLE 2. Approximation qualities based on three MG-DTRS models.

ID	OMG-DTRS	δ-MGDTRS with different values of δ										PMG-DTRS
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
1	1.0000	1.0000	0.9955	0.9856	0.9758	0.9619	0.9215	0.8159	0.6531	0.4589	0.4395	0.4395
2	0.9477	0.9477	0.9159	0.6469	0.6407	0.5260	0.2817	0.2046	0.0533	0.0408	0.0220	0.0220
3	1.0000	1.0000	1.0000	1.0000	1.0000	0.9706	0.7814	0.5628	0.0874	0	0	0
4	0.6449	0.6449	0.2440	0	0	0	0	0	0	0	0	0
5	0.9973	0.9973	0.9744	0.9264	0.9247	0.9092	0	0	0	0	0	0
6	0.6718	0.6718	0.6512	0.5384	0.3256	0.2457	0.1602	0.0005	0	0	0	0
7	0.6274	0.6274	0.6274	0.6274	0.6274	0.6274	0.6274	0.6274	0.6274	0.5712	0.0879	0.0879
8	1.0000	1.0000	1.0000	0.9564	0.8119	0.6931	0.4059	0.3168	0.2475	0	0	0

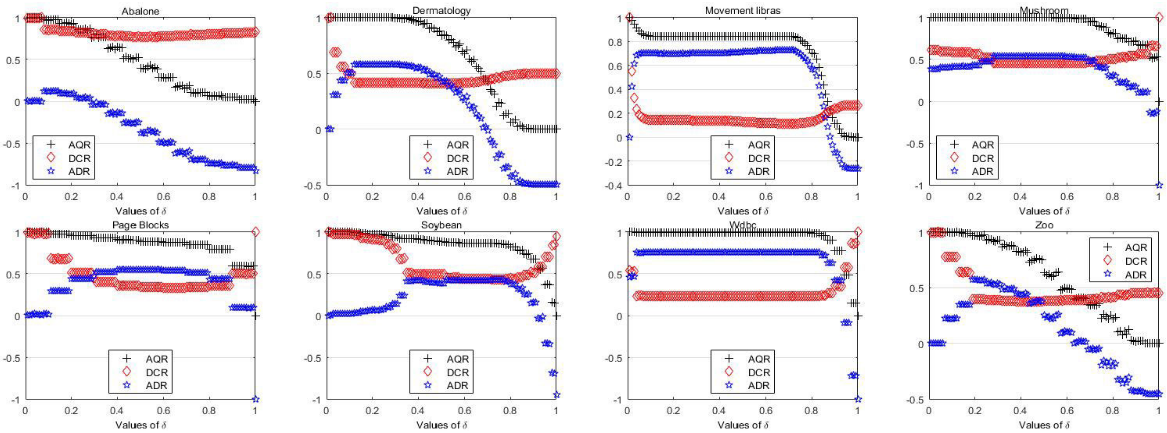

FIGURE 1. Learned threshold by naive approach on several data sets.

B. EXPERIMENTAL ANALYSIS OF THRESHOLD LEARNING ALGORITHM

In this part, the efficiency of the naive approach to learning threshold δ will be presented. In this experiment, a hundred values of δ have already been chosen from 0.01 to 1, meanwhile, for a certain δ and a certain information granule, we generate 100 groups of loss functions in order to calculate thresholds α and β. Learned results by naive approach on 8 data sets are summarized in Fig. 1. Through Fig. 1, we are able to observe:

- Roughly, with the expanding of δ, the values of AQR are diminish. Different from the performance of AQR, with the increasing of δ, the values of DCR are decrease firstly and then increase when it reached a special point.
- The values of ADR in all data sets are regularly change. with the expanding of δ, the values of ADR expand firstly, after reached the peak, the values of ADR are decrease.

Based on the results in Fig. 1, one can obtain the satisfactory δ of all data sets, it can be concluded in Table 3. In Fig. 1, we can note that the satisfactory δ is not only the one in some data sets. For example, in Dermatology, the values of δ from 0.12 to 0.23 all hold the satisfactory δ. In this case, we select an average value of such interval.

C. EXPERIMENTAL ANALYSIS OF INFORMATION GRANULE REDUCTIONS

In the section, several significant kinds of decision rules have been researched. Decision rule figure of reduct is able to be viewed as an evaluation index of data granulation reduct.

Let S be a MDS, $\delta \in (0, 1]$, suppose that $\mathcal{B} \subseteq \mathcal{AT}$ is a reduct of S, then the difference of several decision rules is able to be expressed as below.

$$PDC = \frac{\sum_{i=1}^{n}\left(|POS_\delta(\mathcal{B}, X_i)| - |POS_\delta(\mathcal{AT}, X_i)|\right)}{|U|}; \quad (29)$$

$$BDC = \frac{\sum_{i=1}^{n}\left(|BND_\delta(\mathcal{B}, X_i)| - |BND_\delta(\mathcal{AT}, X_i)|\right)}{|U|}; \quad (30)$$

$$NDC = \frac{\sum_{i=1}^{n}\left(|NEG_\delta(\mathcal{B}, X_i)| - |NEG_\delta(\mathcal{AT}, X_i)|\right)}{|U|}. \quad (31)$$

And then, the values of δ in each data set are selected according to the results in Table 3. The specific experimental consequences are given in Tables 4-9.

Tables 4 to 6 show the decision comparisons among these reducts. In Table 4, the AQ is the approximation quality of original data set. By an investigation of these tables, several conclusions can be obtained.

- Considering the alteration of decision rules, it can be found that the PDC and NDC values based on DM reduct are positive. The BDC value based on DM reduction is negative. The results indicate the reduct based on data mining can increase the number of deterministic rules and reduce the number of vague rules.
- From the viewpoint of reduction algorithms, the function of genetic algorithm is greater than that of heuristic algorithm. For example, in Abalone data set, the PDC

TABLE 3. The satisfactory thresholds of all data sets.

Data ID	1	2	3	4	5	6	7	8
Satisfactory ADR	0.4718	0.1115	0.5828	0.3724	0.5513	0.3097	0.7612	0.5746
Satisfactory δ	0.48	0.11	0.175	0.06	0.41	0.2	0.83	0.24

TABLE 4. The ***PDC*** comparisons of reducts.

Data set	AQ	DP-HAIGR	DM-HAIGR	DM-GAIGR	CM-HAIGR	CM-GAIGR
Abalone	0.9689	0	0.0081	**0.0311**	**0.0311**	-0.1831
Adult4	0.9092	0	0.0855	**0.0874**	0.0630	-0.1265
Dermatology	0.9604	0	**0.0396**	**0.0396**	-0.2178	-0.1287
Molecular Biology	0.7713	0	**0.2013**	0.1804	0.0378	-0.1393
Page Blocks	0.9048	0	0.0347	**0.0663**	0.0347	-0.0274
Steel Plates Faults	0.6497	0	0.0036	**0.0211**	-0.6497	0.0206
Wdbc	0.6257	0	**0.1107**	0.0018	0.1107	0.0070
Zoo	0.9109	0	0.0693	**0.0792**	-0.2673	-0.0396

$^{\star}PDC$ is the difference of positive decision rules between a reduct and original data set, AQ is the basic value and the others are comparison values. The maximal value of PDC is **bolded**.

TABLE 5. The ***BDC*** comparisons of reducts.

Data set	DP-HAIGR	DM-HAIGR	DM-GAIGR	CM-HAIGR	CM-GAIGR
Abalone	0	-0.0206	<u>-0.0436</u>	-0.0280	**0.1343**
Adult4	0	-0.1194	<u>-0.1217</u>	-0.0630	**0.1265**
Dermatology	**0.1188**	<u>-0.1782</u>	<u>-0.1782</u>	0.0792	0.0099
Molecular Biology	0	<u>-0.8264</u>	-0.6086	**0.1063**	-0.0845
Page Blocks	0	-0.0503	<u>-0.1263</u>	**0.0347**	-0.0274
Steel Plates Faults	0	-0.0397	<u>-0.0793</u>	-0.3143	**0.0371**
Wdbc	0	<u>-0.0844</u>	-0.0071	<u>-0.0844</u>	**0.0053**
Zoo	0	<u>-0.3960</u>	0	0	**0.1287**

* BDC is the difference of boundary decision rules between a reduct and original data set. The minimal value of BDC is <u>underlined</u>, meanwhile, the maximal value is **bolded**.

TABLE 6. The ***NDC*** comparisons of reducts.

Data set	DP-HAIGR	DM-HAIGR	DM-GAIGR	CM-HAIGR	CM-GAIGR
Abalone	0	0.0124	0.0124	-0.0031	**0.0488**
Adult4	0	0.0339	**0.0343**	0	0
Dermatology	-0.1188	**0.1386**	**0.1386**	**0.1386**	0.1188
Molecular Biology	0	**0.6200**	0.4282	-0.1441	0.2238
Page Blocks	0	0.0206	**0.0605**	-0.0206	0.0108
Steel Plates Faults	0	0.0361	0.0582	**0.9639**	-0.0577
Wdbc	0	**0.1951**	0.0053	**0.1951**	0.0018
Zoo	0	**0.3267**	0.0792	0.2673	-0.0891

* NDC is the difference of negative decision rules between a reduct and original data set, the maximal value of NDC is **bolded**.

TABLE 7. The ***DCE*** comparisons of reducts.

Data set	Original	DP-HAIGR	DM-HAIGR	DM-GAIGR	CM-HAIGR	CM-GAIGR
Abalone	3524.7	3012.6	972.18	1025.8	490.00	<u>202.78</u>
Adult4	15964	13642	10737	12183	1186.8	<u>760.4</u>
Dermatology	477.38	251.56	251.48	247.24	7.28	<u>7.12</u>
Molecular Biology	27789	23444	11504	13241	<u>244.95</u>	6862.5
Page Blocks	4079.6	3472.9	424.86	1086.8	424.86	<u>248.28</u>
Steel Plates Faults	17569	16771	2147	8202	432.13	<u>320.77</u>
Wdbc	540.79	512.6	61.16	152.95	<u>33.6</u>	42.06
Zoo	311.14	271.48	192.63	162.50	19.26	<u>9</u>

* DCE is the difference of decision cost between a reduct and original data set. The minimal value of DCE is <u>underlined</u>.

value of DM-HAIGR is 0.0081, meanwhile, the *PDC* value of DM-GAIGR is 0.0311, which is much lager than 0.0081.

Table 7 shows the decision cost comparisons among these reducts. Through an investigation of Table 7, we can observe that:

TABLE 8. The *IGLF* comparisons of reducts.

Data set	Original	DP-HAIGR	DM-HAIGR	DM-GAIGR	CM-HAIGR	CM-GAIGR
Abalone	8	**6**	2	2	1	1
Adult4	14	**12**	8	9	1	2
Dermatology	16	**13**	5	5	1	1
Molecular Biology	60	**51**	11	16	1	17
Page Blocks	10	**8**	1	2	1	1
Steel Plates Faults	33	**28**	1	10	1	3
Wdbc	30	**27**	1	12	1	6
Zoo	16	**14**	4	8	1	1

* $IGLF$ is the difference of information granule length between a reduct and original data set. The minimal value of $IGLF$ is underlined, The maximum $IGLF$ of reducts is **bolded**.

TABLE 9. The running time comparisons of reducts (s).

Data set	DP-HAIGR	DM-HAIGR	DM-GAIGR	CM-HAIGR	CM-GAIGR
Abalone	0.0353	0.0316	1.9515	0.0217	2.7885
Adult4	0.1679	0.1777	4.7091	0.0695	5.1897
Dermatology	0.0315	0.0422	1.5839	0.0088	1.7606
Molecular Biology	1.1065	1.2441	2.9688	0.2168	6.1267
Page Blocks	0.0596	0.0326	2.3152	0.0350	2.6856
Steel Plates Faults	0.3729	0.1166	2.6409	0.1070	3.5423
Wdbc	0.2077	0.0320	1.9111	0.0375	2.6240
Zoo	0.0675	0.0798	3.2659	0.0177	1.7760

* The minimal running time of reduct is underlined.

- Comparing with the raw costs, the decision cost of all reduct have been reduced. In terms of the performances of different reducts, we can note that CM based reduct can acquire the min decision costs.
- By comparing the performances of these two reduction approach performs better than genetic approach in several data sets, such as "Wdbc" data set, in some data sets, the genetic approach performs better than heuristic approach, such as "Zoo" data set.

The information granule length factor is a variable to measure the number of information granule elements. If $\mathcal{B} \subseteq \mathcal{AT}$ is the reduct of S, so information granule length factor is able to be expressed as:

$$IGLF = |\mathcal{B}|. \qquad (32)$$

Table 8 shows the information granule length comparisons among different reducts. By a further investigation of Table 8, it is easy to draw the following remarks:

- In Table 8, one can see very clear that the reduct obtained by CM-HAIGR holds only one information granule. This is because in the process of CM-HAIGR, we put the fittest information granule into the reduct in step 4, it is hard to introduce another information granule into the reduct so that the total decision cost can be decreased.
- By comparing different reduction criteria, the DP-HAIGR based reducts hold the most information granule. In other words, in order to maintain the original decision,one needs to hold more information granules, which also indicates the shortage of decision preservation based reduction in MG-DTRS from the perspective of experiments.
- By comparing the performances of these two reduction algorithms, one can observe that the reducts of genetic approach hold more information granules than the ones of heuristic approach.

Finally, Table 9 shows the running time comparisons among these reducts. It is simple to discover that heuristic methods are always faster than genetic methods.

To sum up:

- Decision preservation based reduction may not suitable in such rough set model. Although it can preserve the positive decision rule unchanged, it has several shortages from two points. (1) The original data set does not hold the maximal number of positive decision rules, hence, the reduct which preserves the positive decision rule unchanged would become meaningless. (2) Comparing with the other two reduction criteria, the reduct based on DP criterion needs more decision cost and information.
- The reduct based on DM criterion or CM criterion performs better in such rough set model. On the one hand, the reduct based on DM criterion not only increases the certainty decision with low decision cost, but also decreases the uncertainty comes from border area; the reduct based on CM criterion obtains minimal decision cost.

VI. CONCLUSION

In this article, a novel and generalized framework of MG-DTRS has been studied, which is referred to as δ-MGDTRS model. Different from the traditional MG-DTRS models, a parameterized operator is used to construct the rough approximation. Optimistic and pessimistic

MG-DTRSs are only special cases of our δ-MGDTRS. Furthermore, a satisfactory MG-DTRS model has been defined with high approximation quality and low decision cost, a naive approach is proposed to learn the satisfactory δ.

As far as information granule reduction is concerned, we have discussed the information granule reductions with three different criteria. In order to compute the reducts, heuristic and genetic approaches have been employed for different requirements respectively. The experimental results mainly indicate that the DP based reduction may not suitable in such DTRS models. On the contrary, the DM and CM based reductions can be the popular researches in MG-DTRS model.

The further research is to set the weight of AQR and DCR of the fusion function during the method of obtaining satisfactory threshold. By further optimizing the weight, the model is further improved and applied to more complex practical problems.

ACKNOWLEDGMENT

(Tao Yin and Xiaojuan Mao are co-first authors.)

REFERENCES

[1] K. Cheng, S. Gao, W. Dong, X. Yang, Q. Wang, and H. Yu, "Boosting label weighted extreme learning machine for classifying multi-label imbalanced data," *Neurocomputing*, vol. 403, pp. 360–370, Aug. 2020.

[2] Y. Guo, E. C. C. Tsang, M. Hu, X. Lin, D. Chen, W. Xu, and B. Sang, "Incremental updating approximations for double-quantitative decision-theoretic rough sets with the variation of objects," *Knowl.-Based Syst.*, vol. 189, Feb. 2020, Art. no. 105082.

[3] B. Huang, W.-Z. Wu, J. Yan, H. Li, and X. Zhou, "Inclusion measure-based multi-granulation decision-theoretic rough sets in multi-scale intuitionistic fuzzy information tables," *Inf. Sci.*, vol. 507, pp. 421–448, Jan. 2020.

[4] X. Jia, W. Liao, Z. Tang, and L. Shang, "Minimum cost attribute reduction in decision-theoretic rough set models," *Inf. Sci.*, vol. 219, pp. 151–167, Jan. 2013.

[5] Z. Jiang, H. Dou, J. Song, P. Wang, X. Yang, and Y. Qian, "Data-guided multi-granularity selector for attribute reduction," *Appl. Intell.*, Sep. 2020, doi: 10.1007/s10489-020-01846-6.

[6] H. Ju, W. Pedrycz, H. Li, W. Ding, X. Yang, and X. Zhou, "Sequential three-way classifier with justifiable granularity," *Knowl.-Based Syst.*, vol. 163, pp. 103–119, Jan. 2019.

[7] H. Ju, H. Li, X. Yang, X. Zhou, and B. Huang, "Cost-sensitive rough set: A multi-granulation approach," *Knowl.-Based Syst.*, vol. 123, pp. 137–153, May 2017.

[8] W. Lei, W. Ma, and B. Sun, "Multigranulation behavioral three-way group decisions under hesitant fuzzy linguistic environment," *Inf. Sci.*, vol. 537, pp. 91–115, Oct. 2020.

[9] X. Li and X. Huang, "A novel three-way investment decisions based on decision-theoretic rough sets with hesitant fuzzy information," *Int. J. Fuzzy Syst.*, vol. 22, no. 8, pp. 2708–2719, Mar. 2020.

[10] Z. Li, N. Xie, D. Huang, and G. Zhang, "A three-way decision method in a hybrid decision information system and its application in medical diagnosis," *Artif. Intell. Rev.*, vol. 53, no. 7, pp. 4707–4736, Oct. 2020.

[11] M. Liang, J. Mi, and T. Feng, "Optimal granulation selection for similarity measure-based multigranulation intuitionistic fuzzy decision-theoretic rough sets," *J. Intell. Fuzzy Syst.*, vol. 36, no. 3, pp. 2495–2509, Mar. 2019.

[12] S. J. Liao, Q. X. Zhu, and Y. H. Qian, "Feature–granularity selection with variable costs for hybrid data," *Soft Comput.*, vol. 23, no. 24, pp. 13105–13126, 2019.

[13] D. Liu, X. Yang, and T. Li, "Three-way decisions: Beyond rough sets and granular computing," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 5, pp. 989–1002, May 2020.

[14] K. Liu, X. Yang, H. Fujita, D. Liu, X. Yang, and Y. Qian, "An efficient selector for multi-granularity attribute reduction," *Inf. Sci.*, vol. 505, pp. 457–472, Dec. 2019.

[15] Y. L. Lv, J. Y. Liang, Y. H. Qian, J. J. Wu, and S. Q. Ji, "Sign fusion of multiple QPNs based on qualitative mutual information," *Int. J. Comput. Sci. Eng.*, vol. 19, no. 1, pp. 36–45, 2019.

[16] Z. Pawlak, "Rough set theory and its applications to data analysis," *Cybern. Syst.*, vol. 29, no. 7, pp. 661–688, Oct. 1998.

[17] Y. H. Qian, C. Y. Dang, J. Y. Liang, and W. Z. Wu, "Partial ordering of information granulations: A further investigation," *Expert Syst. J. Knowl. Eng.*, vol. 29, no. 1, pp. 3–24, 2012.

[18] Y. Qian, J. Liang, W. Pedrycz, and C. Dang, "Positive approximation: An accelerator for attribute reduction in rough set theory," *Artif. Intell.*, vol. 174, nos. 9–10, pp. 597–618, Jun. 2010.

[19] Y. Qian, J. Liang, and C. Dang, "Incomplete multigranulation rough set," *IEEE Trans. Syst., Man, Cybern. - Part A: Syst. Humans*, vol. 40, no. 2, pp. 420–431, Mar. 2010.

[20] Y. Qian, J. Liang, Y. Yao, and C. Dang, "MGRS: A multi-granulation rough set," *Inf. Sci.*, vol. 180, no. 6, pp. 949–970, Mar. 2010.

[21] Y. Qian, H. Zhang, Y. Sang, and J. Liang, "Multigranulation decision-theoretic rough sets," *Int. J. Approx. Reasoning*, vol. 55, no. 1, pp. 225–237, Jan. 2014.

[22] J. Qian, C. Liu, D. Miao, and X. Yue, "Sequential three-way decisions via multi-granularity," *Inf. Sci.*, vol. 507, pp. 606–629, Jan. 2020.

[23] N. Rehman, A. Ali, and K. Hila, "Soft dominance based multigranulation decision theoretic rough sets and their applications in conflict problems," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 6079–6110, Dec. 2020.

[24] M.-W. Shao, M.-M. Lv, K.-W. Li, and C.-Z. Wang, "The construction of attribute (object)-oriented multi-granularity concept lattices," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 5, pp. 1017–1032, May 2020.

[25] L. Sun, L. Wang, W. Ding, Y. Qian, and J. Xu, "Neighborhood multi-granulation rough sets-based attribute reduction using lebesgue and entropy measures in incomplete neighborhood decision systems," *Knowl.-Based Syst.*, vol. 192, Mar. 2020, Art. no. 105373.

[26] M. L. Suo, L. F. Tao, B. L. Zhu, X. W. Miao, Z. C. Liang, Y. Ding, X. L. Zhang, and T. Zhang, "Single-parameter decision-theoretic rough set," *Inf. Sci.*, vol. 539, pp. 49–80, Oct. 2020.

[27] A. Tan, W.-Z. Wu, Y. Qian, J. Liang, J. Chen, and J. Li, "Intuitionistic fuzzy rough set-based granular structures and attribute subset selection," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 3, pp. 527–539, Mar. 2019.

[28] G. Tang, F. Chiclana, and P. Liu, "A decision-theoretic rough set model with q-rung orthopair fuzzy information and its application in stock investment evaluation," *Appl. Soft Comput.*, vol. 91, Jun. 2020, Art. no. 106212.

[29] C. Wang, Y. Wang, M. Shao, Y. Qian, and D. Chen, "Fuzzy rough attribute reduction for categorical data," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 818–830, May 2020.

[30] B. Wang and J. Liang, "A novel preference measure for multi-granularity probabilistic linguistic term sets and its applications in large-scale group decision-making," *Int. J. Fuzzy Syst.*, vol. 22, no. 7, pp. 2350–2368, Oct. 2020.

[31] X. Yang, Y. Zhang, H. Fujita, D. Liu, and T. Li, "Local temporal-spatial multi-granularity learning for sequential three-way granular computing," *Inf. Sci.*, vol. 541, pp. 75–97, Dec. 2020.

[32] L. Y. Wen, C. G. Luo, W. Z. Wu, and M. Fan, "Multi-label symbolic value partitioning through random walks," *Neurocomputing*, vol. 387, pp. 195–207, Apr. 2011.

[33] Y. Yao, "Three-way granular computing, rough sets, and formal concept analysis," *Int. J. Approx. Reasoning*, vol. 116, pp. 106–125, Jan. 2020.

[34] Y. Y. Yao, "Probabilistic approaches to rough sets," *Expert Syst.*, vol. 20, no. 5, pp. 287–297, Nov. 2003.

[35] Y. Yao, "The superiority of three-way decisions in probabilistic rough set models," *Inf. Sci.*, vol. 181, no. 6, pp. 1080–1096, Mar. 2011.

[36] Y. Yao, "Two semantic issues in a probabilistic rough set model," *Fundamenta Informaticae*, vol. 108, nos. 3–4, pp. 249–265, 2011.

[37] J. Zhan, H. Jiang, and Y. Yao, "Covering-based variable precision fuzzy rough sets with PROMETHEE-EDAS methods," *Inf. Sci.*, vol. 538, pp. 314–336, Oct. 2020.

[38] H.-Y. Zhang and S.-Y. Yang, "Three-way group decisions with interval-valued decision-theoretic rough sets based on aggregating inclusion measures," *Int. J. Approx. Reasoning*, vol. 110, pp. 31–45, Jul. 2019.

[39] C. Zhang, D. Li, and J. Liang, "Multi-granularity three-way decisions with adjustable hesitant fuzzy linguistic multigranulation decision-theoretic rough sets over two universes," *Inf. Sci.*, vol. 507, pp. 665–683, Jan. 2020.

[40] C. Zhang, D. Li, and J. Liang, "Interval-valued hesitant fuzzy multi-granularity three-way decisions in consensus processes with applications to multi-attribute group decision making," *Inf. Sci.*, vol. 511, pp. 192–211, Feb. 2020.

TAO YIN received the B.Sc. degree in computer science and technology from the Jiangsu Institute of Technology, Changzhou, China, in 2020. He is currently pursuing the master's degree with the School of Information Science and Technology, Nantong University. His main research interests include rough sets and granular computing.

XIAOJUAN MAO received the B.Sc. and M.Sc. degrees in the clinical major of integrated traditional chinese and western medicine from the Zhejiang University of Traditional Chinese Medicine, Hangzhou, China, in 2013 and 2016, respectively. She is currently working with The Sixth People's Hospital of Nantong. Her main research interests include respiratory diseases and knowledge discovery.

YING ZHANG is currently pursuing the bachelor's degree with the School of Information Science and Technology, Nantong University. Her current research interests include rough sets and big data analysis.

YITING MA is currently pursuing the bachelor's degree with the School of Information Science and Technology, Nantong University. Her current research interests include rough sets and big data analysis.

HENGRONG JU received the B.Sc. and M.Sc. degrees in computer science and technology from the Jiangsu University of Science and Technology, Zhenjiang, China, in 2012 and 2015, respectively, and the Ph.D. degree in management science and engineering from Nanjing University, Nanjing, China, in 2019. From 2017 to 2018, he has worked as a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is currently a Lecturer with the School of Information Science and Technology, Nantong University. He has authored or coauthored more than 20 scientific papers in international journals and conferences. His current research interests include knowledge discovery and granular computing.

WEIPING DING (Senior Member, IEEE) received the Ph.D. degree in computation application from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2013. He was a Visiting Scholar with the University of Lethbridge (UL), Alberta, Canada, in 2011. From 2014 to 2015, he was a Postdoctoral Researcher with the Brain Research Center, National Chiao Tung University (NCTU), Hsinchu, Taiwan. In 2016, he was a Visiting Scholar with the National University of Singapore (NUS), Singapore. From 2017 to 2018, he was a Visiting Professor with the University of Technology Sydney (UTS), Ultimo, NSW, Australia. He has published more than 50 papers in flagship journals and conference proceedings as the first author, including IEEE Transactions on Fuzzy Systems, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Cybernetics, and so on. To data, he has held 12 approved invention patents in total more than 20 issued patents. His main research interests include granular computing data mining, machine learning, and big data analytics. He serves/served as a Program Committee Member for seven international conferences and workshops. He is a member of Task Force on Adaptive and Evolving Fuzzy Systems of IEEE CIS FSTC, a member of Soft Computing TC of IEEE SMC, a member of Granular Computing (GrC) TC of IEEE SMC, and also a member of Data Mining and Big Data Analytics TC of IEEE CIS. He currently serves on the Editorial Advisory Board for *Knowledge-Based Systems*. He serves as an Associate Editor for IEEE Transactions on Fuzzy Systems, *Information Sciences*, *Swarm and Evolutionary Computation*, *Journal of Intelligent and Fuzzy Systems*, and IEEE Access, and a leading Guest Editor for four international journals.