Article

Cost-Sensitive Multigranulation Approximation in Decision-Making Applications

Jie Yang [1,2], **Juncheng Kuang** [2], **Qun Liu** [2] **and Yanmin Liu** [1,*]

[1] School of Physics and Electronic Science, Zunyi Normal University, Zunyi 563002, China
[2] Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[*] Correspondence: yanmin7813@163.com; Tel.: +86-1508-606-1907

Abstract: A multigranulation rough set (MGRS) model is an expansion of the Pawlak rough set, in which the uncertain concept is characterized by optimistic and pessimistic upper/lower approximate boundaries, respectively. However, there is a lack of approximate descriptions of uncertain concepts by existing information granules in MGRS. The approximation sets of rough sets presented by Zhang provide a way to approximately describe knowledge by using existing information granules. Based on the approximation set theory, this paper proposes the cost-sensitive multigranulation approximation of rough sets, i.e., optimistic approximation and pessimistic approximation. Their related properties were further analyzed. Furthermore, a cost-sensitive selection algorithm to optimize the multigranulation approximation was performed. The experimental results show that when multigranulation approximation sets and upper/lower approximation sets are applied to decision-making environments, multigranulation approximation produces the least misclassification costs on each dataset. In particular, misclassification costs are reduced by more than 50% at each granularity on some datasets.

Keywords: multigranulation rough sets; optimistic approximation; pessimistic approximation; cost-sensitive; decision-making applications

1. Introduction

As a human-inspired paradigm, granular computing (GrC) solves complex problems by utilizing multiple granular layers [1–4]. Zadeh [1] noted that information in granules refer to pieces, classes, and groups, into which complex information is divided in accordance with the characteristics and processes of understanding and decision-making. From the different views, GrC models mainly cover four types: fuzzy sets [5], rough sets [6], quotient spaces [7], and cloud models [8]. As representative models of GrC, rough sets describe uncertain concepts by upper and lower approximation boundaries, which have been applied to data mining [9,10], medical systems [11], attribute reductions [12,13], decision systems [14,15], and machine learning [16].

Regarding similarity, Zhang [17–20] presented the approximation set of rough sets, vague sets, rough fuzzy sets, rough vague sets, etc. These approximation models were developed by utilizing the existing equivalence classes to describe uncertain concepts. The approximation model has a higher similarity with the target concept than the upper/lower approximations. Furthermore, the approximation model has been applied in attribute reduction [21], image segmentation [22], the optimization algorithm [23], etc. Based on the approximation set theory, Yang [24,25] developed the approximation model of rough sets based on misclassification costs. In the process of cost-sensitive learning, the smaller misclassification costs will help to improve the decision-making qualities in real applications. Recently, from the perspective of three-way decisions [26–29], Yao [30] constructed a symbol–meaning–value (SMV) model for data analysis. In the three-way

decision model, the equivalence classes in a boundary region will produce misclassification costs when they are used as approximation sets. Hence, the approximation model that is constructed from the perspective of similarity is no longer applicable to cost-sensitive scenarios. To minimize the misclassification costs of constructing the approximation set, we proposed the multigranulation approximation, i.e., the optimistic approximation model and pessimistic approximation model. Moreover, to search the optimal approximation layer for multigranulation rough sets [31] under the constraints, the algorithm of the cost-sensitive multigranulation approximation selection is further proposed to be applied to decision-making environments.

The following sections are arranged as follows: Section 2 presents the related works. Section 3 introduces the relevant definitions of the multigranulation rough set and approximation set. Section 4 introduces an approximate representation of the rough sets. Section 5 presents the cost-sensitive multigranulation approximations of rough sets and further introduces the optimal multigranulation approximation algorithm. To verify the availability of the proposed model, the related experiments and discussion are presented in Section 6. Ultimately, in Section 7, the conclusions are presented.

2. Literature Review

Rough sets are typically constructed based on a single binary relation. However, in many cases, they may be described in multiple granularity structures. In order to extend single granularity to multi-granularity in rough approximations, Qian [31] proposed the multigranulation rough set model (MGRS), where the upper/lower approximations were defined by multi-equivalence relations (multiple granulations) in the universe [32,33]. For the lower approximation of optimistic MGRS, at least one granular space was obtained, such that objects completely belonged to the target concept. For the lower approximation of pessimistic MGRS, objects completely belong to target concepts in each granular space. MGRS has two advantages: (1) In the process of decision-making applications, the decision of each decision maker may be independent of the same project (or an element) in the universe [34]. In this situation, the intersection operations between any two granularity structures will be redundant for decision-making [35]. (2) Extract decision rules from distributive information systems and groups of intelligent agents by using rough set approaches [34,36].

There are many works [33–35,37–42] on multigranulation rough sets. To extend the MGRS to the neighborhood information system, Hu [43,44] presented matrix-based incremental approaches to update knowledge about neighborhood information systems by changing the granular structures. From the perspective of uncertainty measure, Sun [39] proposed a feature selection based on fuzzy neighborhood multigranulation rough sets. Xu [38] proposed a dynamic approximation update mechanism of a multigranulation neighborhood rough set from a local viewpoint. Liu [35] introduced a parameter-free multigranularity attribute reduction scheme, which is more effective for microarray data than other well-established attribute reductions. Based on the three-way decision theory, She [40] presented a five-valued logic approach for the multigranulation rough set model. From the above, however, the method of approximately describing the target concept with existing information granules is not given, which limits the application of the multigranulation rough set theory. Li [41] presented two kinds of local multigranulation rough set models in the ordered decision system by extending the single granulation environment to a multigranulation case. Zhang [42] constructed hesitant fuzzy multigranulation rough sets to handle the hesitant fuzzy information and group decision-making for person–job fit.

3. Preliminaries

In this section, some necessary definitions related to the multigranulation rough set and approximation set are reviewed to facilitate the framework of this paper. Let $S = (U, C \cup D, V, f)$ be a decision information table, where U is a non-empty finite set of

objects, C is a non-empty finite set of condition attributes, D is a set of decision attributes, V is the set of all attribute values, and $f : U \times C$ is an information function.

Definition 1 ((Rough Sets) [6]). *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A \subseteq C$ and $X \subseteq U$, the lower and upper approximation sets of X are given as follows:*

$$\underline{A}(X) = \{\, x \in U | [x]_A \subseteq X \},$$
$$\overline{A}(X) = \{\, x \in U | [x]_A \cap X \neq \phi \}.$$

where $[x]_A$ denotes the equivalence class induced by U/A, namely, $U/A = \{[x]_R\} = \{[x]_1, [x]_2, \cdots, [x]_l\}$.

Based on the lower and upper approximations, the universe U can be divided into three disjoint regions, which are expressed as follows:

$$POS_A(X) = \underline{A}(X),$$
$$BND_A(X) = \overline{A}(X) - \underline{A}(X),$$
$$NEG_A(X) = U - \overline{A}(X).$$

Definition 2 ((Optimistic multigranulation rough sets) [36]). *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$ and $X \subseteq U$, then the lower and upper approximation sets of X related to $A_1, A_2, \ldots, A_m$ are given as follows:*

$$\underline{\sum_{i=1}^{m} A_i^O}(X) = \{\, x | [x]_{A_1} \subseteq X \vee [x]_{A_2} \subseteq X \vee \ldots \vee [x]_{A_m} \subseteq X, x \in U \}, \tag{1}$$

$$\overline{\sum_{i=1}^{m} A_i^O}(X) = \sim \underline{\sum_{i=1}^{m} A_i^O}(\sim X). \tag{2}$$

Then, $(\underline{\sum_{i=1}^{m} A_i^O}(X), \overline{\sum_{i=1}^{m} A_i^O}(X))$ is called optimistic multigranulation rough sets. The lower and upper approximation sets of X in optimistic multigranulation rough sets are presented by multiple independent approximation spaces. The boundary regions are defined as follows:

$$BND^O_{\sum_{i=1}^{m} A_i}(X) = \overline{\sum_{i=1}^{m} A_i^O}(X) - \underline{\sum_{i=1}^{m} A_i^O}(X). \tag{3}$$

Definition 3 ((Pessimistic multigranulation rough sets) [36]). *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$, and $X \subseteq U$. The lower and upper approximation sets of X related to $A_1, A_2, \ldots, A_m$ are given as follows:*

$$\underline{\sum_{i=1}^{m} A_i^P}(X) = \{\, x | [x]_{A_1} \subseteq X \wedge [x]_{A_2} \subseteq X \wedge \ldots \wedge [x]_{A_m} \subseteq X, x \in U \}, \tag{4}$$

$$\overline{\sum_{i=1}^{m} A_i^P}(X) = \sim \underline{\sum_{i=1}^{m} A_i^P}(\sim X). \tag{5}$$

Then, $(\underline{\sum_{i=1}^{m} A_i^P}(X), \overline{\sum_{i=1}^{m} A_i^P}(X))$ is called pessimistic multigranulation rough sets. The lower and upper approximation sets of X in pessimistic multigranulation rough sets are presented by

multiple independent approximation spaces. However, the strategy is different from optimistic multigranulation rough sets. The boundary region is defined as follows:

$$BND^P_{\sum\limits_{i=1}^{m} A_i}(X) = \overline{\sum_{i=1}^{m} A_i^P(X)} - \underline{\sum_{i=1}^{m} A_i^P(X)}. \tag{6}$$

Definition 4 ((Approximation of rough sets) [17]). *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A \subseteq C$ and $X \subseteq U$. $U/A = \{[x]_1, [x]_2, \cdots, [x]_l\}$ is a granularity layer on U, then the α-approximation of X on U/A is defined as follows:*

$$A_\alpha(X) = \cup\{[x]_i | \overline{\mu}([x]_i) \geq \alpha, [x]_i \subseteq U\}. \tag{7}$$

where $0 \leq \alpha \leq 1$, $\overline{\mu}([x]_i) = \frac{|[x]_i \cap X|}{|[x]_i|}$ denotes the membership degree of the equivalence class $[x]_i$ belongs to X.

Example 1. *Let Table 1 be a decision information table, $A_1, A_2, A_3 \subseteq C$ and $X \subseteq U$. For the element x_4, the equivalence classes $[x_4]_i (i = 1, 2, 3)$ belonging to X in the multigranulation approximation space are as follows:*

$$[x_4]_1 = \{x_1, x_2, x_3, x_4\};$$
$$[x_4]_2 = \{x_3, x_4, x_7, x_8\};$$
$$[x_4]_3 = \{x_2, x_4, x_6, x_8\}.$$

Accordingly, the membership degrees are computed:

$$\overline{\mu}([x_4]_1) = \frac{0+0+0+1}{4} = 0.25;$$
$$\overline{\mu}([x_4]_2) = \frac{0+1+1+1}{4} = 0.75;$$
$$\overline{\mu}([x_4]_3) = \frac{0+1+1+1}{4} = 0.75.$$

Table 1. A decision information table.

	A_1	A_2	A_3	X
x_1	0	0	0	0
x_2	0	0	1	0
x_3	0	1	0	0
x_4	0	1	1	1
x_5	1	0	0	0
x_6	1	0	1	1
x_7	1	1	0	1
x_8	1	1	1	1

If α is set to 0.5, considering the optimistic approximation, element x_4 will be classified into the optimistic lower approximation sets of X due to one of its membership degrees being greater than 0.5. However, if considering the pessimistic approximation, element x_4 will only be classified into the pessimistic upper approximation sets of X.

Based on the given conditions, we have:

$$X = \frac{0.25 + 0.25 + 0.25}{x_1} + \frac{0.25 + 0.25 + 0.75}{x_2} + \frac{0.25 + 0.75 + 0.05}{x_3} + \frac{0.25 + 0.75 + 0.75}{x_4}$$
$$+ \frac{0.75 + 0.25 + 0.25}{x_5} + \frac{0.75 + 0.25 + 0.75}{x_6} + \frac{0.75 + 0.75 + 0.25}{x_7} + \frac{0.75 + 0.75 + 0.75}{x_8}.$$

Then, the results of the optimistic approximations are shown as follows:

$$\overline{\sum_{i=1}^{m} A_i^O(X)} = \{x_2, x_3, x_4, x_5, x_6, x_7, x_8\},$$

$$\underline{\sum_{i=1}^{m} A_i^O(X)} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\},$$

$$BND_{\sum_{i=1}^{m} A_i}^O(X) = \{x_1\}.$$

Moreover, the results of the pessimistic approximations, in this case, are changed as follows:

$$\overline{\sum_{i=1}^{m} A_i^P(X)} = \{x_8\},$$

$$\underline{\sum_{i=1}^{m} A_i^P(X)} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\},$$

$$BND_{\sum_{i=1}^{m} A_i}^P(X) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}.$$

4. Cost-Sensitive Approximation Methods of the Rough Sets

Let $S = (U, C \cup D, V, f)$ be a decision information table, $A \subseteq C$ and $X \subseteq U$. $U/A = \{[x]_1, [x]_2, \cdots, [x]_l\}$ is a granularity layer on U. λ_{12} represents the cost generated by taking an element belonging to X as the approximation, λ_{21} represents the cost generated by taking an element that does not belong to X as the approximation. Furthermore, misclassification costs incurred by the equivalence classes in characterizing X on U/A are given in the following:

$$\lambda^Y = \lambda_{12}(1 - \overline{\mu}([x]_i))|[x]_i|. \tag{8}$$

Misclassification costs incurred by the equivalence classes when not characterizing X on U/A are given in the following:

$$\lambda^N = \lambda_{21}\overline{\mu}([x]_i)|[x]_i|. \tag{9}$$

Herein, $\overline{\mu}([x]_i)$ $(i = 1, 2, \ldots, l)$ denotes the membership degree of $[x]_i$ belonging to X. Based on the Bayesian decision procedure, the minimum cost decision rules are expressed as follows:

(P) If $\lambda^Y \leq \lambda^N$, $[x]_i \subseteq A(X)$;
(N) If $\lambda^Y > \lambda^N$, $[x]_i \not\subseteq A(X)$.

It is clear that the above rules are only relevant to the loss function $\overline{\mu}([x]_i)$. From Formulas (8) and (9), the decision rules are re-expressed in the following:

(P1) If $\overline{\mu}([x]_i) \geq \frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}}$, $[x]_i \subseteq A(X)$;
(N1) If $\overline{\mu}([x]_i) < \frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}}$, $[x]_i \not\subseteq A(X)$.

Supposing $\gamma = \frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}}$, then we have the following decision rules:

(P2) If $\overline{\mu}([x]_i) \geq \gamma$, $[x]_i \subseteq A(X)$;
(N2) If $\overline{\mu}([x]_i) < \gamma$, $[x]_i \not\subseteq A(X)$.

Definition 5. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A \subseteq C$ and $X \subseteq U$. $U/A = \{[x]_1, [x]_2, \cdots, [x]_l\}$ is a granularity layer on U, then the cost-sensitive approximation model (CSA) of rough sets is defined as follows:*

$$A(X) = \cup\left\{[x]_i \middle| \overline{\mu}([x]_i) \geq \frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}}, [x]_i \subseteq U\right\}. \tag{10}$$

Suppose $0 \leq \lambda_{12}, \lambda_{21} \leq 1$, boundary region I and boundary region II are denoted by $BN1(X) = \{ [x]_i | \frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}} \leq \overline{\mu}([x]_i) < 1 \}$ and $BN2(X) = \{ [x]_i | 0 < \overline{\mu}([x]_i) < \frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}} \}$, respectively, then $BN(X) = BN1(X) \cup BN2(X)$, and $A(X) = BN1(X) \cup POS(X)$. Figure 1 shows the CSA of rough sets, where $BN1(X)$ is the dark blue region, which denotes the region in the boundary region used as the approximation. $BN2(X)$ is the light blue region, which denotes the region in the boundary region not used as the approximation. Therefore, the region surrounded by the green broken line in Figure 1 constructs the approximations of rough sets, and the misclassification costs come from two uncertain regions, defined as follows:

$$DC(A(X)) = \sum_{[x]_i \in BN1(X)} \lambda^Y + \sum_{[x]_i \in BN2(X)} \lambda^N. \tag{11}$$
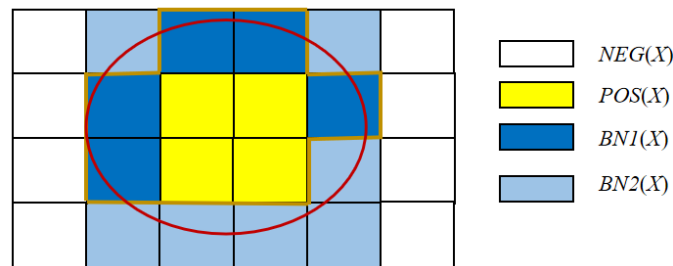

Figure 1. The approximation of rough sets (surrounded by the broken line).

Theorem 1. *Let $S = (U, A \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$, then $DC(A_1(X)) \geq DC(A_2(X))$.*

Proof of Theorem 1. Let U be a non-empty finite domain, $U/A_1 = \{E_1, E_2, \cdots, E_l\}$ and $U/A_2 = \{F_1, F_2, \cdots, F_m\}$. Because $A_1 \subseteq A_2$, $U/A_2 \preccurlyeq U/A_1$. According to the condition, for simplicity, supposing only one granule E_1 can be subdivided into two finer sub-granules by $\Delta A = A_2 - A_1$ (the more complicated cases can be transformed into this case, so we will not repeat them here). Without loss of generality, let $E_1 = F_1 \cup F_2$, $E_2 = F_3$, $E_3 = F_4, \ldots, E_l = F_m$ ($m = l + 1$), namely, $U/A_2 = \{F_1, F_2, E_2, E_3, \ldots, E_l\}$. There are two cases to prove the theorem as follows:

(1) Suppose $\overline{\mu}(E_1) \geq \gamma$, obviously, $E_1 \subseteq A_1(X)$.

Case 1. $\overline{\mu}(F_1) \geq \gamma$ and $\overline{\mu}(F_2) \geq \gamma$. Namely, $F_1 \subseteq A(X)$ and $F_2 \subseteq A(X)$. Case 1, in which the granules are subdivided in $BN1(X)$, can be shown in Figure 2a, then

$$
\begin{aligned}
\Delta DC_{A_1 - A_2}(X) &= DC(A_1(X)) - DC(A_2(X)) \\
&= \lambda_{12}(1 - \overline{\mu}(E_1))|E_1| - \lambda_{12}(1 - \overline{\mu}(F_1))|F_1| - \lambda_{12}(1 - \overline{\mu}(F_2))|F_2|. \\
&= (|E_1| - |F_1| - |F_2| + \sum_{x_i \in F_1} \mu(x_i) + \sum_{x_i \in F_2} \mu(x_i) - \sum_{x_i \in E_1} \mu(x_i))\lambda_{12}.
\end{aligned}
$$

Because $\sum_{x_i \in E_1} \mu(x_i) = \sum_{x_i \in F_1} \mu(x_i) + \sum_{x_i \in F_2} \mu(x_i)$ and $|E_1| = |F_1| + |F_2|$, then $\Delta DC_{A_1 - A_2} = 0$. Therefore, $DC(A_1(X)) = DC(A_2(X))$.

Case 2. $\overline{\mu}(F_1) \geq \gamma$ and $\overline{\mu}(F_2) < \gamma$. Namely, $F_1 \subseteq A(X)$ and $F_2 \not\subseteq A(X)$. Case 2, in which the granules are subdivided in $BN1(X)$, can be shown in Figure 2b, then

$$
\begin{aligned}
\Delta DC_{A_1 - A_2} &= DC(A_1(X)) - DC(A_2(X)) \\
&= \lambda_{12}(1 - \overline{\mu}(E_1))|E_1| - \lambda_{12}(1 - \overline{\mu}(F_1))|F_1| - \lambda_{21}\overline{\mu}(F_2)|F_2| \\
&= |F_2|(\lambda_{12} - \overline{\mu}(F_2)(\lambda_{21} + \lambda_{12}))
\end{aligned}
$$

Because $\overline{\mu}(F_2) < \gamma = \frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}}$ and $|E_1| = |F_1| + |F_2|$, then $\Delta DC_{A_1 - A_2}(X) \geq 0$. Therefore, $DC(A_1(X)) > DC(A_2(X))$.

(2)　Suppose $\bar{\mu}(E_1) < \gamma$, obviously, $E_1 \not\subset A_1(X)$.

Case 1. $\bar{\mu}(F_1) \geq \gamma$ and $\bar{\mu}(F_2) < \gamma$. Namely, $F_1 \subseteq A(X)$ and $F_2 \not\subset A(X)$. Case 1, in which the granules are subdivided in $BN2(X)$, can be shown in Figure 2c, then

$$\begin{aligned}
\Delta DC_{A_1 - A_2} &= DC(A_1(X)) - DC(A_2(X)) \\
&= \bar{\mu}(E_1)|E_1|\lambda_{21} - \bar{\mu}(F_2)|F_2|\lambda_{21} - (1 - \bar{\mu}(F_1))|F_1|\lambda_{12} \\
&= |F_1|\bar{\mu}(F_1)((\lambda_{21} + \lambda_{12}) - \lambda_{12}).
\end{aligned}$$

Because $\bar{\mu}(F_1) \geq \gamma = \frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}}$, then $\Delta DC_{A_1 - A_2} \geq 0$. Therefore, $DC(A_1(X)) \geq DC(A_2(X))$.

Case 2. $\bar{\mu}(F_1) < \gamma$ and $\bar{\mu}(F_2) < \gamma$. Namely, $F_1 \not\subset R(X)$ and $F_2 \not\subset R(X)$. Case 2, in which the granules are subdivided in $BN2(X)$, can be shown in Figure 2d, then

$$\begin{aligned}
\Delta DC_{A_1 - A_2} &= DC(A_1(X)) - DC(A_2(X)) \\
&= \bar{\mu}(E_1)|E_1|\lambda_{21} - \bar{\mu}(F_1)|F_1|\lambda_{21} - \bar{\mu}(F_2)|F_2|\lambda_{21} \\
&= (\sum_{x_i \in E_1} \mu(x_i) - \sum_{x_i \in F_1} \mu(x_i) - \sum_{x_i \in F_2} \mu(x_i))\lambda_{21}.
\end{aligned}$$

(a)

(b)

(c)

(d)
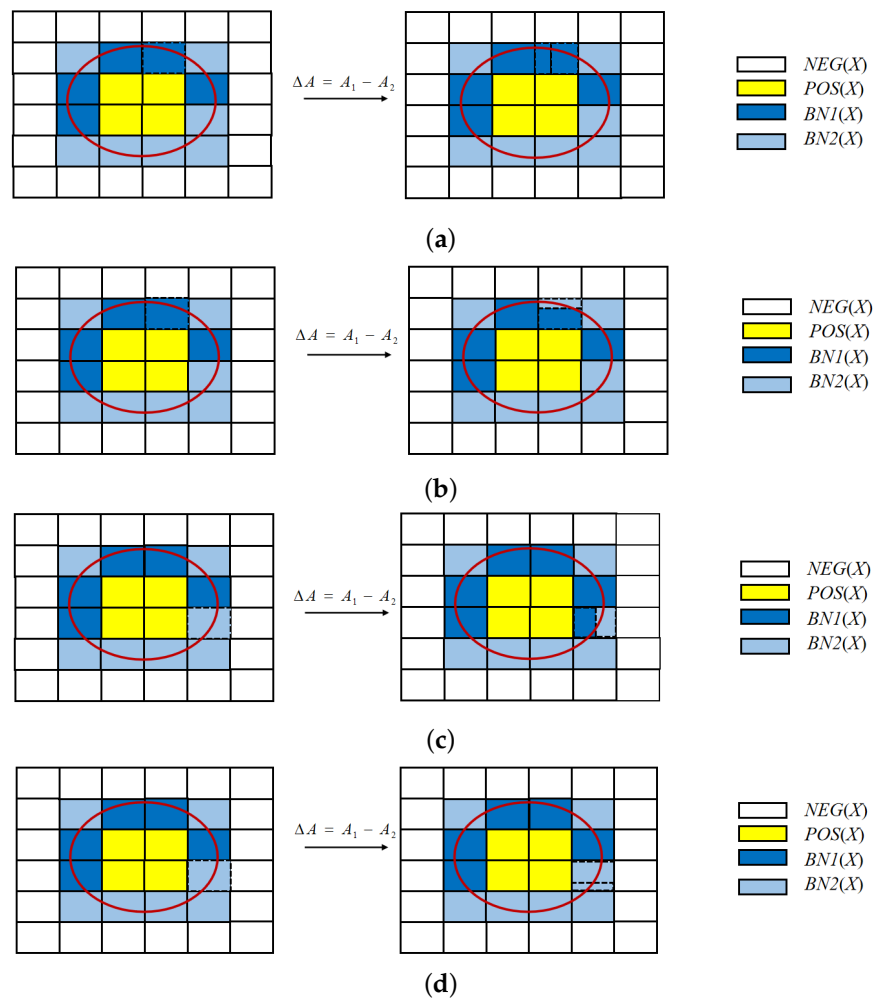
Figure 2. The granules subdivided in $BN1(X)$ and $BN2(X)$ of the cost-sensitive approximation model of rough sets. And all the red circles in the figure represent the set X. In addition, (**a**) shows the case 1 in which the granules are subdivided in $BN1(X)$; (**b**) shows the case 2 in which the granules are subdivided in $BN1(X)$; (**c**) shows the case 1 in which the granules are subdivided in $BN2(X)$; (**d**) shows the case 2 in which the granules are subdivided in $BN2(X)$.

Because $\sum\limits_{x_i \in E_1} \mu(x_i) = \sum\limits_{x_i \in F_1} \mu(x_i) + \sum\limits_{x_i \in F_2} \mu(x_i)$, then $\Delta DC_{A_1 - A_2} = 0$. Therefore, $DC(A_1(X)) = DC(A_2(X))$. □

Theorem 1 shows that the misclassification costs in the approximation model will monotonously decrease with the changing approximation space, in accordance with human cognitive mechanisms.

5. Cost-Sensitive Multigranulation Approximations and Optimal Granularity Selection Method

The multigranulation rough set model (MGRS) [43] extends single granularity to multi-granularity in rough approximations to describe an uncertain concept. MGRS is an expansion of the classical rough set, and the target concept is characterized by optimistic and pessimistic upper/lower approximation boundaries in MGRS, respectively. However, there is a lack of an approximate description of an uncertain concept by utilizing the equivalence classes in MGRS. In this section, based on the model proposed in Section 3, we further construct the approximations of MGRS.

5.1. Cost-Sensitive Multigranulation Approximations of Rough Sets

Definition 6. *Let* $S = (U, C \cup D, V, f)$ *be a decision information table,* $A_1, A_2, \ldots, A_m \subseteq C$ *and* $X \subseteq U$. *The optimistic membership degree of* $x \in U$ *related to* $A_1, A_2, \ldots, A_m$ *is given as follows:*

$$\overline{\mu}^O_{\sum\limits_{i=1}^{m} A_i}(x) = \max\{\overline{\mu}([x]_{A_i}) | i = 1, 2, \ldots, m\}. \tag{12}$$

Definition 7. *Let* $S = (U, C \cup D, V, f)$ *be a decision information table,* $A_1, A_2, \ldots, A_m \subseteq C$ *and* $X \subseteq U$. *The approximation model of the optimistic MGRS of X related to* $A_1, A_2, \ldots, A_m$ *is given as follows:*

$$\sum_{i=1}^{m} A_i^O(X) = \{ x | \overline{\mu}([x]_{A_1}) \geq \gamma \vee \overline{\mu}([x]_{A_2}) \geq \gamma \vee \ldots \vee \overline{\mu}([x]_{A_m}) \geq \gamma, x \in U \}. \tag{13}$$

From the perspective of the optimistic membership degree, $\sum\limits_{i=1}^{m} A_i^O(X)$ *can be =expressed as follows:*

$$\sum_{i=1}^{m} A_i^O(X) = \{ x | \overline{\mu}^O_{\sum\limits_{i=1}^{m} A_i}(x) \geq \gamma, x \in U \}. \tag{14}$$

The corresponding decision regions are defined as follows:

$$BN1^O_{\sum\limits_{i=1}^{m} A_i}(X) = \{ x | 1 > \overline{\mu}^O_{\sum\limits_{i=1}^{m} A_i}(x) \geq \gamma, x \in U \},$$

$$BN2^O_{\sum\limits_{i=1}^{m} A_i}(X) = \{ x | 0 < \overline{\mu}^O_{\sum\limits_{i=1}^{m} A_i}(x) < \gamma, x \in U \},$$

$$POS^O_{\sum\limits_{i=1}^{m} A_i}(X) = \{ x | \overline{\mu}^O_{\sum\limits_{i=1}^{m} A_i}(x) = 1, x \in U \},$$

$$NEG^O_{\sum\limits_{i=1}^{m} A_i}(X) = \{ x | \overline{\mu}^O_{\sum\limits_{i=1}^{m} A_i}(x) = 0, x \in U \}.$$

We have

$$\sum_{i=1}^{m} A_i^O(X) = BN1^O_{\sum\limits_{i=1}^{m} A_i}(X) \cup POS^O_{\sum\limits_{i=1}^{m} A_i}(X). \tag{15}$$

The misclassification costs of approximations of optimistic MGRS come from two uncertain regions $BN1^O_{\sum\limits_{i=1}^{m} A_i}(X)$ and $BN2^O_{\sum\limits_{i=1}^{m} A_i}(X)$, which are defined in the following:

$$DC(\sum_{i=1}^{m} A_i^O(X)) = \sum_{x \in BN1^O_{\sum\limits_{i=1}^{m} A_i}(X)} \lambda^Y + \sum_{x \in BN2^O_{\sum\limits_{i=1}^{m} A_i}(X)} \lambda^N. \tag{16}$$

Theorem 2. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$, $X \subseteq U$ and $X_1, X_2, \ldots, X_n \subseteq U$. The following properties hold:*

(1) $\sum\limits_{i=1}^{m} A_i^O(X) = \bigcup_{i=1}^{m} A_i(X)$;

(2) $\sum\limits_{i=1}^{m} A_i^O(\bigcap_{j=1}^{n} X_j) = \bigcup_{i=1}^{m} (\bigcap_{j=1}^{n} A_i^O(X_j))$;

(3) $\sum\limits_{i=1}^{m} A_i^O(\bigcap_{j=1}^{n} X_j) \subseteq \bigcap_{j=1}^{n} (\sum\limits_{i=1}^{m} A_i^O(X_j))$;

(4) $\sum\limits_{i=1}^{m} A_i^O(\bigcup_{j=1}^{n} X_j) \supseteq \bigcup_{j=1}^{n} (\sum\limits_{i=1}^{m} A_i^O(X_j))$;

(5) $\underline{\sum\limits_{i=1}^{m} A_i^O(X)} \subseteq \sum\limits_{i=1}^{m} A_i^O(X) \subseteq \overline{\sum\limits_{i=1}^{m} A_i^O}(X)$.

Proof of Theorem 2.

(1) From Formula (14), $\sum\limits_{i=1}^{m} A_i^O(X) = \bigcup_{i=1}^{m} A_i(X)$ obviously holds.

(2) $\sum\limits_{i=1}^{m} A_i^O(\bigcap_{j=1}^{n} X_j) = \bigcup_{i=1}^{m} A_i^O(\bigcap_{j=1}^{n} X_j) = \bigcup_{i=1}^{m} (\bigcap_{j=1}^{n} A_i^O(X_j))$.

(3) $\sum\limits_{i=1}^{m} A_i^O(\bigcap_{j=1}^{n} X_j) = \bigcup_{i=1}^{m} (\bigcap_{j=1}^{n} A_i(X_j))$

$= \bigcap_{j=1}^{n} (\bigcup_{i=1}^{m} A_i(X_j) \cap \cdots$

$= \bigcap_{j=1}^{n} (\sum\limits_{i=1}^{m} A_i^O(X_j)) \cap \cdots$

$= \bigcap_{j=1}^{n} (\sum\limits_{i=1}^{m} A_i^O(X_j))$.

(4) From $X_j \subseteq \bigcup_{j=1}^{n} X_j$, we have $\sum\limits_{i=1}^{m} A_i^O(X_j) \subseteq \sum\limits_{i=1}^{m} A_i^O(\bigcup_{j=1}^{n} X_j)$. Therefore, $\sum\limits_{i=1}^{m} A_i^O(\bigcup_{j=1}^{n} X_j) \supseteq \bigcup_{j=1}^{n} (\sum\limits_{i=1}^{m} A_i^O(X_j))$.

(5) It is easy to prove by Formulas (1), (2), and (14). $\square$

Definition 8. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$ and $X \subseteq U$. The pessimistic membership degree of $x \in U$ related to $A_1, A_2, \ldots, A_m$ is given as follows:*

$$\overline{\mu}^P_{\sum\limits_{i=1}^{m} A_i}(x) = \min\{\overline{\mu}([x]_{A_i}) | i = 1, 2, \ldots, m\}. \tag{17}$$

Definition 9. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$ and $X \subseteq U$. The approximation model of pessimistic MGRS of X related to $A_1, A_2, \ldots, A_m$ is given as follows:*

$$\sum_{i=1}^{m} A_i^P(X) = \{x | \overline{\mu}([x]_{A_1}) \geq \gamma \wedge \overline{\mu}([x]_{A_2}) \geq \gamma \wedge \ldots \wedge \overline{\mu}([x]_{A_m}) \geq \gamma, x \in U\}. \tag{18}$$

From the perspective of the pessimistic membership degree, $\sum_{i=1}^{m} A_i^P(X)$ can be expressed as follows:

$$\sum_{i=1}^{m} A_i^P(X) = \{x | \overline{\mu}_{\sum_{i=1}^{m} A_i}^P(x) \geq \gamma, x \in U\}. \tag{19}$$

The corresponding decision regions are expressed in the following:

$$BN1_{\sum_{i=1}^{m} A_i}^P(X) = \{x | 1 > \overline{\mu}_{\sum_{i=1}^{m} A_i}^P(x) \geq \gamma, x \in U\},$$

$$BN2_{\sum_{i=1}^{m} A_i}^P(X) = \{x | 0 < \overline{\mu}_{\sum_{i=1}^{m} A_i}^P(x) < \gamma, x \in U\},$$

$$POS_{\sum_{i=1}^{m} A_i}^P(X) = \{x | \overline{\mu}_{\sum_{i=1}^{m} A_i}^P(x) = 1, x \in U\},$$

$$NEG_{\sum_{i=1}^{m} A_i}^P(X) = \{x | \overline{\mu}_{\sum_{i=1}^{m} A_i}^P(x) = 0, x \in U\}.$$

We have

$$\sum_{i=1}^{m} A_i^P(X) = BN1_{\sum_{i=1}^{m} A_i}^P(X) \cup POS_{\sum_{i=1}^{m} A_i}^P(X). \tag{20}$$

The misclassification costs of approximations of optimistic MGRS come from two uncertain regions $BN1_{\sum_{i=1}^{m} A_i}^P(X)$ and $BN2_{\sum_{i=1}^{m} A_i}^P(X)$, which are defined in the following:

$$DC(\sum_{i=1}^{m} A_i^P(X)) = \sum_{x \in BN1_{\sum_{i=1}^{m} A_i}^P(X)} \lambda^Y + \sum_{x \in BN1_{\sum_{i=1}^{m} A_i}^P(X)} \lambda^N. \tag{21}$$

Theorem 3. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$, $X \subseteq U$ and $X_1, X_2, \ldots, X_n \subseteq U$. The following properties hold:*

(1) $\sum\limits_{i=1}^{m} A_i^P(X) = \bigcap_{i=1}^{m} (A_i(X))$;

(2) $\sum\limits_{i=1}^{m} A_i^P(\bigcap_{j=1}^{n} X_j) = \bigcap_{j=1}^{n} (\sum\limits_{i=1}^{m} A_i^P(X_j))$;

(3) $\sum\limits_{i=1}^{m} A_i^P(\bigcup_{j=1}^{n} X_j) \supseteq \bigcup_{j=1}^{n} (\sum\limits_{i=1}^{m} A_i^P(X_j))$;

(4) $\underline{\sum\limits_{i=1}^{m} A_i^P(X)} \subseteq \sum\limits_{i=1}^{m} A_i^P(X) \subseteq \overline{\sum\limits_{i=1}^{m} A_i^P(X)}$.

Proof of Theorem 3.

(1) $\forall x \in \sum\limits_{i=1}^{m} A_i^P(X)$, according to Definition 9, $\overline{\mu}([x]_{A_i}) \geq \gamma$ holds, $i = 1, 2, \cdots, m$. According to Definition 5, $x \in A_i(X)$, $i = 1, 2, \cdots, m$ and $x \in \bigcap_{i=1}^{m} (A_i(X))$ holds, i.e., $\sum\limits_{i=1}^{m} A_i^P(X) \subseteq \bigcap_{i=1}^{m} (A_i(X))$. $\forall x \in \bigcap_{i=1}^{m} (A_i(X))$, $\overline{\mu}([x]_{A_i}) \geq \gamma$, $i = 1, 2, \cdots, m$. According to Definition 7, $x \in \sum\limits_{i=1}^{m} A_i^P(X)$ holds, i.e., $\bigcap_{i=1}^{m} (A_i(X)) \subseteq \sum\limits_{i=1}^{m} A_i^P(X)$. Therefore, we have $\sum\limits_{i=1}^{m} A_i^P(X) = \bigcap_{i=1}^{m} (A_i(X))$.

(2) From the proof of (1), $\sum_{i=1}^{m} A_i^P(\bigcap_{j=1}^{n} X_j) = \bigcap_{i=1}^{m} A_i(\bigcap_{j=1}^{n} X_j)$. According to Definition 7, $\bigcap_{i=1}^{m} A_i(\bigcap_{j=1}^{n} X_j) = \bigcap_{i=1}^{m} \bigcap_{j=1}^{n} A_i(X_j)$. Because $\bigcap_{i=1}^{m} A_i(X_j) = \sum_{i=1}^{m} A_i^P(X_j)$,

$$\sum_{i=1}^{m} A_i^P(\bigcap_{j=1}^{n} X_j) = \bigcap_{i=1}^{m} A_i(\bigcap_{j=1}^{n} X_j) = \bigcap_{j=1}^{n} \bigcap_{i=1}^{m} A_i(X_j) = \bigcap_{j=1}^{n} (\sum_{i=1}^{m} A_i^P(X_j)).$$

(3) $\forall x \in \bigcup_{j=1}^{n} (\sum_{i=1}^{m} A_i^P(X_j)), \exists X_k (k \in \{1, 2, \cdots, n\}), x \in \sum_{i=1}^{m} A_i^P(X_j)$. According to Definition 7, $\forall X_j, j = 1, 2, \cdots, n, \overline{\mu}([x]_{A_i}) \geq \gamma, , i = 1, 2, \cdots, m, x \in \sum_{i=1}^{m} A_i^P(\bigcup_{j=1}^{n} (X_j))$ holds.

Therefore, $\sum_{i=1}^{m} A_i^P(\bigcup_{j=1}^{n} X_j) \supseteq \bigcup_{j=1}^{n} (\sum_{i=1}^{m} A_i^P(X_j))$.

(4) It is easy to prove by Formulas (4), (5), and (19).
$\square$

Theorem 4. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$ and $X \subseteq U$, $A = A_1 \cup A_2 \cup \ldots \cup A_m$ and $X_1, X_2, \ldots, X_n \subseteq U$. The following properties hold:*

(1) $\sum_{i=1}^{m} A_i^P(X) \subseteq \sum_{i=1}^{m} A_i^O(X) \subseteq A(X)$;

(2) $\sum_{i=1}^{m-1} A_i^O(X) \subseteq \sum_{i=1}^{m} A_i^O(X)$ *and* $\sum_{i=1}^{m-1} A_i^P(X) \supseteq \sum_{i=1}^{m} A_i^P(X)$.

Proof of Theorem 4.

(1) According to Definition 6, we only need to prove $\sum_{i=1}^{m} A_i^P(X) \subseteq \sum_{i=1}^{m} A_i^O(X)$. $\forall x \in \sum_{i=1}^{m} A_i^P(X)$, according to Definition 7, $\overline{\mu}([x]_{A_i}) \geq \gamma$. From Definition 5, we have $x \in \sum_{i=1}^{m} A_i^O(X)$.

(2) It is easy to prove according to Definitions 5 and 7.
$\square$

Lemma 1. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A \subseteq C$ and $X \subseteq U$, $U/A = \{E_1, E_2, \ldots, E_l\}$ is a granularity layer on U. The following properties hold:*

(1) $\sum_{E_i \in BN1(X)} \lambda_{E_i}^Y \leq \sum_{E_i \in BN1(X)} \lambda_{E_i}^N$;

(2) $\sum_{E_i \in BN2(X)} \lambda_{E_i}^N \leq \sum_{E_i \in BN2(X)} \lambda_{E_i}^Y$. *Here, $E_i \in U/A$ ($i = 1, 2, \ldots, l$).*

Proof of Lemma 1.

(1) $\lambda_{E_i}^Y - \lambda_{E_i}^N = \lambda_{12}(1 - \overline{\mu}(E_i))|E_i| - \lambda_{21}\overline{\mu}(E_i)|E_i| = |E_i|(\lambda_{12} - (\lambda_{12} + \lambda_{21}))\overline{\mu}(E_i)$, because $E_i \in BN1(X)$, we have $\frac{\lambda_{12}}{\lambda_{12}+\lambda_{21}} \leq \overline{\mu}(E_i) < 1$, then $\lambda_{E_i}^Y \leq \lambda_{E_i}^N$, therefore $\sum_{E_i \in BN1(X)} \lambda_{E_i}^Y \leq \sum_{E_i \in BN1(X)} \lambda_{E_i}^N$.

(2) $\lambda_{E_i}^Y - \lambda_{E_i}^N = \lambda_{12}(1 - \overline{\mu}(E_i))|E_i| - \lambda_{21}\overline{\mu}(E_i)|E_i| = |E_i|(\lambda_{12} - (\lambda_{12} + \lambda_{21}))\overline{\mu}(E_i)$. Because $E_i \in BN2(X)$, we have $0 < \overline{\mu}(E_i) < \frac{\lambda_{12}}{\lambda_{12}+\lambda_{21}}$, then $\lambda_{E_i}^Y \geq \lambda_{E_i}^N$. Therefore, $\sum_{E_i \in BN2(X)} \lambda_{E_i}^N \leq \sum_{E_i \in BN2(X)} \lambda_{E_i}^Y$.
$\square$

Lemma 1 shows that the misclassification costs incurred by the equivalence classes in characterizing X are not more than the misclassification costs incurred by the equivalence classes when they do not characterize X in $BN1(X)$. Moreover, misclassification costs incurred by the equivalence classes when they do not characterize X are not more than the misclassification costs incurred by the equivalence classes in characterizing X in $BN2(X)$.

Theorem 5. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A \subseteq C$ and $X \subseteq U$, the following properties hold:*

(1) $\displaystyle\sum_{x \in BN1^O_{\sum_{i=1}^m A_i}(X)} \lambda^Y \leq \sum_{x \in BN1^O_{\sum_{i=1}^m A_i}(X)} \lambda^N$ and $\displaystyle\sum_{x \in BN1^P_{\sum_{i=1}^m A_i}(X)} \lambda^Y \leq \sum_{x \in BN1^P_{\sum_{i=1}^m A_i}(X)} \lambda^N$;

(2) $\displaystyle\sum_{x \in BN2^O_{\sum_{i=1}^m A_i}(X)} \lambda^N \leq \sum_{x \in BN2^O_{\sum_{i=1}^m A_i}(X)} \lambda^Y$ and $\displaystyle\sum_{x \in BN1^P_{\sum_{i=1}^m A_i}(X)} \lambda^N \leq \sum_{x \in BN1^P_{\sum_{i=1}^m A_i}(X)} \lambda^Y$.

Proof of Theorem 5. From Lemma 1, Theorem 5 holds. $\square$

Theorem 5 shows that misclassification costs incurred by the equivalence classes in characterizing X are not more than misclassification costs incurred by the equivalence classes when they do not characterize X in $BN1^O_{\sum_{i=1}^m A_i}(X)$ and $BN1^P_{\sum_{i=1}^m R_i}(X)$. Moreover, misclassification costs incurred by the equivalence classes when they do not characterize X are not more than misclassification costs incurred by the equivalence classes in characterizing X in $BN2^O_{\sum_{i=1}^m A_i}(X)$ and $BN2^P_{\sum_{i=1}^m A_i}(X)$.

$DC(\sum_{i=1}^m A_i^O(X))$, $DC(\underline{\sum_{i=1}^m A_i^O(X)})$ and $DC(\overline{\sum_{i=1}^m A_i^O(X)})$ denote the misclassification costs generated when $\sum_{i=1}^m A_i^O(X)$, $\underline{\sum_{i=1}^m A_i^O(X)}$ and $\overline{\sum_{i=1}^m A_i^O(X)}$ are approximated to X, respectively. Then, the following theorem holds:

Theorem 6. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A \subseteq C$ and $X \subseteq U$, Then,*
$DC(\sum_{i=1}^m A_i^O(X)) \leq DC(\underline{\sum_{i=1}^m A_i^O(X)})$ and $DC(\sum_{i=1}^m A_i^O(X)) \leq DC(\overline{\sum_{i=1}^m A_i^O(X)})$.

Proof of Theorem 6. When $\underline{\sum_{i=1}^m A_i^O(X)}$ is taken as the approximation of X, $DC(\underline{\sum_{i=1}^m A_i^O(X)}) = \sum_{x \in BN(X)} \lambda^N$; when $\overline{\sum_{i=1}^m A_i^O(X)}$ is considered the approximation of X, $DC(\overline{\sum_{i=1}^m A_i^O(X)}) = \sum_{x \in BN(X)} \lambda^Y$; when $DC(\sum_{i=1}^m A_i^O(X))$ is considered the approximation of X, $DC(\sum_{i=1}^m A_i^O(X)) = \sum_{x \in BN1(X)} \lambda^Y + \sum_{x \in BN2(X)} \lambda^N$.

Because $BN(X) = BN1(X) \cup BN2(X)$, we have:

$$DC(\underline{\sum_{i=1}^m A_i^O(X)}) = \sum_{x \in BN1(X)} \lambda^N + \sum_{x \in BN2(X)} \lambda^N,$$

$$DC(\overline{\sum_{i=1}^m A_i^O(X)}) = \sum_{x \in BN1(X)} \lambda^Y + \sum_{x \in BN2(X)} \lambda^Y.$$

Therefore, according to Theorem 5,

$$DC(\sum_{i=1}^m A_i^O(X)) \leq DC(\underline{\sum_{i=1}^m A_i^O(X)}),$$

$$DC(\sum_{i=1}^m A_i^O(X)) \leq DC(\overline{\sum_{i=1}^m A_i^O(X)}).$$

Theorem 6 indicates that when $DC(\sum\limits_{i=1}^{m} A_i^O(X))$, $DC(\underline{\sum\limits_{i=1}^{m} A_i^O(X)})$ and $DC(\overline{\sum\limits_{i=1}^{m} A_i^O}(X))$ are used as approximations of X, respectively, $DC(\sum\limits_{i=1}^{m} A_i^O(X))$ generates the least misclassification costs. $\square$

$DC(\sum\limits_{i=1}^{m} A_i^P(X))$, $DC(\underline{\sum\limits_{i=1}^{m} A_i^P(X)})$ and $DC(\overline{\sum\limits_{i=1}^{m} A_i^P}(X))$ denote the misclassification costs generated when $\sum\limits_{i=1}^{m} A_i^P(X)$, $\underline{\sum\limits_{i=1}^{m} A_i^P(X)}$ and $\overline{\sum\limits_{i=1}^{m} A_i^P}(X)$ are approximated to X, respectively.

Theorem 7. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A \subseteq C$ and $X \subseteq U$. Then, $DC(\sum\limits_{i=1}^{m} A_i^P(X)) \leq DC(\underline{\sum\limits_{i=1}^{m} A_i^P(X)})$ and $DC(\sum\limits_{i=1}^{m} A_i^P(X)) \leq DC(\overline{\sum\limits_{i=1}^{m} A_i^P}(X))$.*

Proof of Theorem 7. Similar to Theorem 6, Theorem 7 is easy to prove. $\square$

From Theorem 7, when $DC(\sum\limits_{i=1}^{m} A_i^P(X))$, $DC(\underline{\sum\limits_{i=1}^{m} A_i^P(X)})$ and $DC(\overline{\sum\limits_{i=1}^{m} A_i^P}(X))$ are used as approximations of X, respectively, $DC(\sum\limits_{i=1}^{m} A_i^P(X))$ generates the least misclassification costs. Theorems 6 and 7 reflect the advantages of the multigranulation approximation sets that are used for approximating the target concept.

5.2. The Optimal Multigranulation Approximation Selection Method

The objects in the boundary region may be reclassified under different granularities. As a result, the equivalence classes used to represent the approximation set will be changed in the boundary region. In practical applications, the optimal approximation selection should consider both the misclassification and test costs. In MGRS, uncertain concepts characterized in a finer approximation layer result in lower misclassification costs, and test costs increase with the added attributes. Therefore, it is essential to find a balance between misclassification and test costs.

Lemma 2. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$, $A_1 \subseteq A_2 \subseteq, \ldots, \subseteq A_m$, and $X \subseteq U$, then $\forall x \in U$, $\overline{\mu}_{\sum\limits_{i=1}^{m-1} A_i}^{O}(x) \leq \overline{\mu}_{\sum\limits_{i=1}^{m} A_i}^{O}(x)$.*

Lemma 3. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$, $A_1 \subseteq A_2 \subseteq, \ldots, \subseteq A_m$, and $X \subseteq U$, then $\forall x \in U$, $\overline{\mu}_{\sum\limits_{i=1}^{m-1} A_i}^{P}(x) \geq \overline{\mu}_{\sum\limits_{i=1}^{m} A_i}^{P}(x)$.*

Theorem 8. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$, $A_1 \subseteq A_2 \subseteq, \ldots, \subseteq A_m$ and $X \subseteq U$, only when $\overline{\mu}(x)(x \in BN1_{\sum\limits_{i=1}^{m-1} A_i}^{O}(X))$ changed with the attribute increased, then $DC(\sum\limits_{i=1}^{m-1} A_i^O(X)) \geq DC(\sum\limits_{i=1}^{m} A_i^O(X))$.*

Proof of Theorem 8.

$$DC(\sum_{i=1}^{m} A_i^O(X)) = \sum_{x \in BN1^O_{\sum_{i=1}^{m} A_i}(X)} \lambda^Y + \sum_{x \in BN2^O_{\sum_{i=1}^{m} A_i}(X)} \lambda^N$$

$$= \sum_{x \in BN1^O_{\sum_{i=1}^{m} A_i}(X)} \lambda_{12}(1 - \overline{\mu}(x)) + \sum_{x \in BN2^O_{\sum_{i=1}^{m} A_i}(X)} \lambda_{21}\overline{\mu}(x).$$

According to Lemma 2, we have

$$\overline{\mu}^O_{\sum_{i=1}^{m-1} A_i}(x) \le \overline{\mu}^O_{\sum_{i=1}^{m} A_i}(x).$$

Obviously, $DC(\sum_{i=1}^{m-1} A_i^O(X)) \ge DC(\sum_{i=1}^{m} A_i^O(X))$. □

From Theorem 7, for optimistic MGRS, to reduce the misclassification costs of the approximation, we can add the attribute that only changes the membership of objects in $x \in BN1^O_{\sum_{i=1}^{m-1} A_i}(X)$.

Theorem 9. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$ and $A_1 \subseteq A_2 \subseteq, \ldots, \subseteq A_m, X \subseteq U$, only when $\overline{\mu}(x)$ ($x \in BN2^P_{\sum_{i=1}^{m-1} A_i}(X)$) changed with the attributes increased, then $DC(\sum_{i=1}^{m-1} A_i^P(X)) \ge DC(\sum_{i=1}^{m} A_i^P(X))$.*

Proof of Theorem 9.

$$DC(\sum_{i=1}^{m} A_i^P(X)) = \sum_{x \in BN1^P_{\sum_{i=1}^{m} A_i}(X)} \lambda^Y + \sum_{x \in BN2^P_{\sum_{i=1}^{m} A_i}(X)} \lambda^N$$

$$= \sum_{x \in BN1^P_{\sum_{i=1}^{m} A_i}(X)} \lambda_{12}(1 - \overline{\mu}(x)) + \sum_{x \in BN2^P_{\sum_{i=1}^{m} A_i}(X)} \lambda_{21}\overline{\mu}(x).$$

According to Lemma 3, we have

$$\overline{\mu}^P_{\sum_{i=1}^{m-1} A_i}(x) \ge \overline{\mu}^P_{\sum_{i=1}^{m} A_i}(x).$$

Obviously, $DC(\sum_{i=1}^{m-1} A_i^P(X)) \ge DC(\sum_{i=1}^{m} A_i^P(X))$. □

From Theorem 9, for pessimistic MGRS, to reduce the misclassification costs of the approximation, we can add the attribute that only changes the membership of objects in $x \in BN2^O_{\sum_{i=1}^{m-1} R_i}(X)$.

In practical applications, on the one hand, the factors included in the test cost, such as money, time, environment, etc., are hard to evaluate objectively. On the other hand, these factors are hard to be integrated because of their different dimensions. In this section, we will evaluate test costs in an attribute-driven form, which are more objective.

Definition 10. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $a \in C$ and $X \subseteq U$, the significance of a is defined as follows:*

$$Sig(a, C, D) = DC_{C-\{a\}} - DC_C. \tag{22}$$

Definition 11. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $a \in C$, $A \subseteq C$ and $X \subseteq U$; the test cost to construct $A(X)$ is defined as follows:*

$$TC_A = \sum_{a \in A} Sig(a, C, D). \tag{23}$$

In this paper, for simplicity, to present the optimal granularity selection of the multigranulation approximation, we only use the optimistic MGRS as an example.

Definition 12. *Let $S = (U, C \cup D, V, f)$ be a decision information table, $A_1, A_2, \ldots, A_m \subseteq C$ and $X \subseteq U$; the test cost to construct $\sum_{i=1}^{m} A_i^O(X)$ can be defined as follows:*

$$TC_{\sum_{i=1}^{m} A_i^O} = \sum_{i=1}^{m} TC_{A_i}. \tag{24}$$

In this paper, the misclassification and test costs for user requirements are represented as DC_u and TC_u, respectively. A multigranulation approximation $\sum_{i=1}^{k} A_i^O(X)$ is selected to meet the constraints $DC_{\sum_{i=1}^{k} A_i^O(X)} \leq DC_u$ and $TC_{\sum_{i=1}^{k} A_i^O(X)} \leq TC_u$, then the related decision are made on $\sum_{i=1}^{k} A_i^O(X)$. Figure 3 presents the optimal multigranulation approximation selection of optimistic MGRS. Herein, $\sum_{i=1}^{3} A_i^O(X)$ complies with the requirements of the misclassification costs and fails to comply with the requirements of the test costs; $A_1^O(X)$ complies with the requirements of the test cost and fails to comply with the requirements of misclassification costs; $\sum_{i=1}^{2} A_i^O(X)$ complies with both requirements of misclassification costs and test costs, enabling effective calculations according to granularity optimization. Similarly, the optimal approximation selection of pessimistic MGRS is the same. We formalize the computation as an optimization problem:

$$\arg\min_{k} Cost_{\sum_{i=1}^{k} A_i^O(X)}, \tag{25}$$

$$s.t.$$

$$\xi DC_{\sum_{i=1}^{k} A_i^O(X)} \leq DC_u;$$

$$TC_{\sum_{i=1}^{k} A_i^O(X)} \leq TC_u.$$

where $Cost_{\sum_{i=1}^{k} A_i^O(X)} = \xi DC_{\sum_{i=1}^{k} A_i^O} + TC_{\sum_{i=1}^{k} A_i^O}$, and $Cost_{\sum_{i=1}^{k} A_i^O(X)}$ denotes the total cost for constructing $\sum_{i=1}^{k} A_i^O(X)$. $\xi = \frac{|U|}{\left| \sum_{i=1}^{m} A_i^O(X) \right|} * \frac{1}{DC(A_m(X))}$ reflects the contribution degree of the multigranulation approximation layer for the misclassification costs of $\sum_{i=1}^{k} A_i^O(X)$.
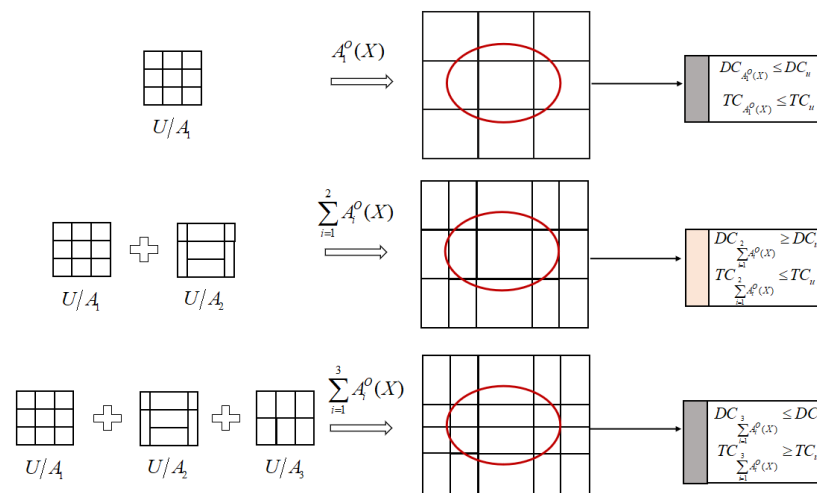
Figure 3. The optimal granularity selection of the optimistic multigranulation approximation. And the red circles in the figure represent the set X.

5.3. Case Study

Table 2 is an evaluation form of the company venture capital given by five experts. $U = \{x_1, x_2, \ldots, x_{900}\}$ is 900 investment plans, which are evaluated by 5 experts. The risk level is divided into three categories, including high and low. Suppose that $DC_u = 5$ and $TC_u = 2.5$.

(1) According to the above conditions, the attribute significance can be computed by Formula (25), which is shown in Table 3.

(2) Each multigranulation approximation is obtained by adding attributes in ascending order of attribute significance, $E_4 \to E_3 \to E_2 \to E_1 \to E_5$ in each stage, respectively. We represent the attributes as follows: $A_1 = E_4$, $A_2 = E_3$, $A_3 = E_2$, $A_4 = E_1$, and $A_5 = E_5$.

(3) For each multigranulation approximation layer, $DC_{\sum_{i=1}^{k} A_i^O}$, $TC_{\sum_{i=1}^{k} A_i^O}$, and $Cost_{\sum_{i=1}^{k} A_i^O(X)}$ are computed by Formulas (16), (24), and (25), respectively, where $k = 1, 2, \ldots, 5$ and the results are displayed in Table 4.

Table 2. Evaluation form of the company's venture capital.

Firm	E_1	E_2	E_3	E_4	E_5	D
x_1	3	3	3	3	1	High
x_2	2	1	2	3	2	High
x_3	2	1	2	1	2	High
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
x_{898}	3	3	2	2	3	Low
x_{899}	3	1	3	3	1	Low
x_{900}	1	1	3	3	1	Low

Table 3. Description of attribute significance.

Attribute	E_1	E_2	E_3	E_4	E_5
$Sig(a, C, D)$	0.74	0.56	0.54	0.36	0.77

Table 4. The description of cost in each multigranulation approximation.

	$A_1^O(X)$	$\sum_{i=1}^{2} A_i^O(X)$	$\sum_{i=1}^{3} A_i^O(X)$	$\sum_{i=1}^{4} A_i^O(X)$	$\sum_{i=1}^{5} A_i^O(X)$
ξDC	8.3	5.9	4.7	3.6	3.5
TC	0.36	0.9	1.46	2.2	2.97
$Cost$	8.66	6.8	6.16	5.8	6.47

$Cost_k$ changes with the increased attributes and only $Cost_{\sum_{i=1}^{3} A_i^O(X)}$ and $Cost_{\sum_{i=1}^{4} A_i^O(X)}$ satisfy $DC_{k_{\sum_{i=1}^{k} A_i^O(X)}} \leq DC_u$ and $TC_{k_{\sum_{i=1}^{k} A_i^O(X)}} \leq TC_u$ at the same time. According to the Formula (25), we choose the multigranulation approximation layer with the lowest total cost from the above layers; its corresponding approximation layer is $\sum_{i=1}^{3} A_i^O(X)$. Therefore, $\sum_{i=1}^{3} A_i^O(X)$ is the optimal multigranulation approximation used for deciding investment plans, because it possesses lower misclassification costs, i.e., from the perspective of optimistic MGRS, E_4, E_3, and E_2 are reasonable expert sets. The analysis of the case study shows that the proposed method can search for a reasonable approximation under the constraint conditions.

6. Simulation Experiment and Result Analysis

6.1. Simulation Experiment

In this section, the effectiveness and rationality of our model are demonstrated as shown by illustrative experiments. The computer used in the experiments was a WIN 10 operating system with 3.10-GHz CPU and 16.0 GB RAM, and the programming software was MATLAB R2022a. The capability of the proposed model was evaluated on six UCI datasets, which are shown in Table 5. In our experiments, we randomly took away some known attribute values from datasets 10–12 to create incomplete decision systems. The missing values are randomly distributed on all conditional attributes.

Table 5. The description of datasets.

ID	Dataset	Attribute Characteristics	Instances	Condition Attributes
1	Bank	Integer	39	12
2	Breast-Cancer	Integer	699	9
3	Car	Integer	1728	6
4	ENB2012data	Real	768	8
5	Mushroom	Integer	8124	22
6	Tic	Integer	958	9
7	Air Quality	Real	9358	12
8	Concrete	Real	1030	8
9	Hcv	Real	569	10
10	Wisconsin	Real	699	9
11	Zoo	Integer	101	16
12	Balance	Integer	625	4

From Figure 4, for classical rough sets, the misclassification costs of the approximation model monotonously decrease with the granularity being finer, which complies with human cognitive habits.
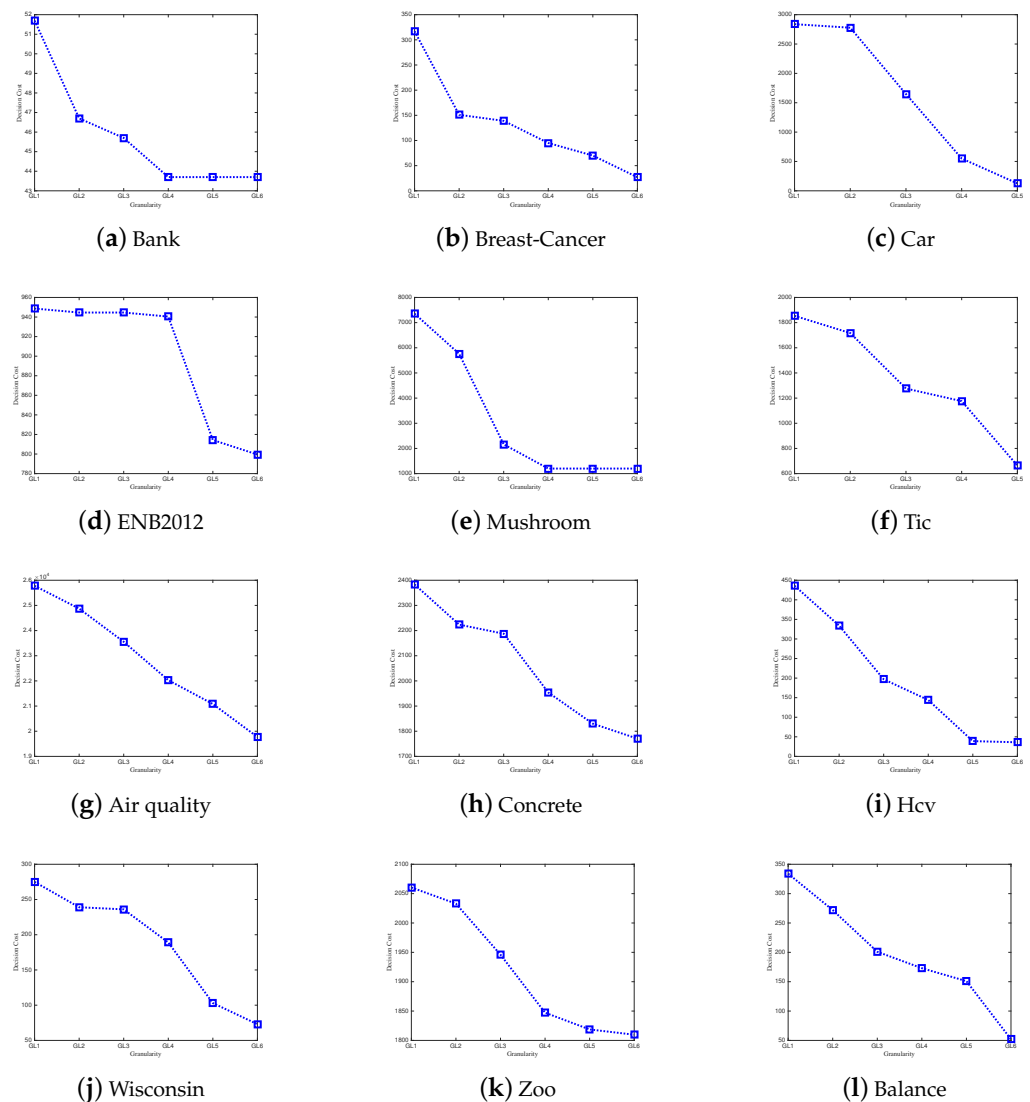
(a) Bank   **(b)** Breast-Cancer   **(c)** Car

(d) ENB2012   **(e)** Mushroom   **(f)** Tic

(g) Air quality   **(h)** Concrete   **(i)** Hcv

(j) Wisconsin   **(k)** Zoo   **(l)** Balance

Figure 4. The misclassification cost with the changing granularity on each dataset.

In Figure 5, O_DC1, O_DC2, O_DC3 and O_DC4 represent $\sum\limits_{x \in BN1^{O_m}_{\sum_{i=1} A_i}(X)} \lambda^Y$, $\sum\limits_{x \in BN1^{O_m}_{\sum_{i=1} A_i}(X)} \lambda^N$, $\sum\limits_{x \in BN2^{O_m}_{\sum_{i=1} A_i}(X)} \lambda^N$ and $\sum\limits_{x \in BN2^{O_m}_{\sum_{i=1} A_i}(X)} \lambda^Y$, respectively. P_DC1, P_DC2, P_DC3 and P_DC4 represent $\sum\limits_{x \in BN1^{P_m}_{\sum_{i=1} A_i}(X)} \lambda^Y$, $\sum\limits_{x \in BN1^{P_m}_{\sum_{i=1} A_i}(X)} \lambda^N$, $\sum\limits_{x \in BN2^{P_m}_{\sum_{i=1} A_i}(X)} \lambda^N$ and $\sum\limits_{x \in BN1^{P_m}_{\sum_{i=1} A_i}(X)} \lambda^Y$, respectively.

From Figure 5, under different granular layers, misclassification costs incurred by the equivalence classes in approximating X are always less than or equal to misclassification costs incurred by the equivalence classes when they do not characterize X in $BN1^{O_m}_{\sum_{i=1} A_i}(X)$ and $BN1^{P_m}_{\sum_{i=1} A_i}(X)$. Moreover, misclassification costs incurred by equivalence classes when they do not characterize X are less than or equal to the misclassification costs incurred by the equivalence classes in approximating X in $BN2^{O_m}_{\sum_{i=1} A_i}(X)$ and $BN2^{P_m}_{\sum_{i=1} A_i}(X)$. This is consistent with Theorem 4.

(a) Bank

(b) Breast-Cancer

(c) Car

(d) ENB2012

(e) Mushroom

(f) Tic

(g) Air quality

(h) Concrete

(i) Hcv

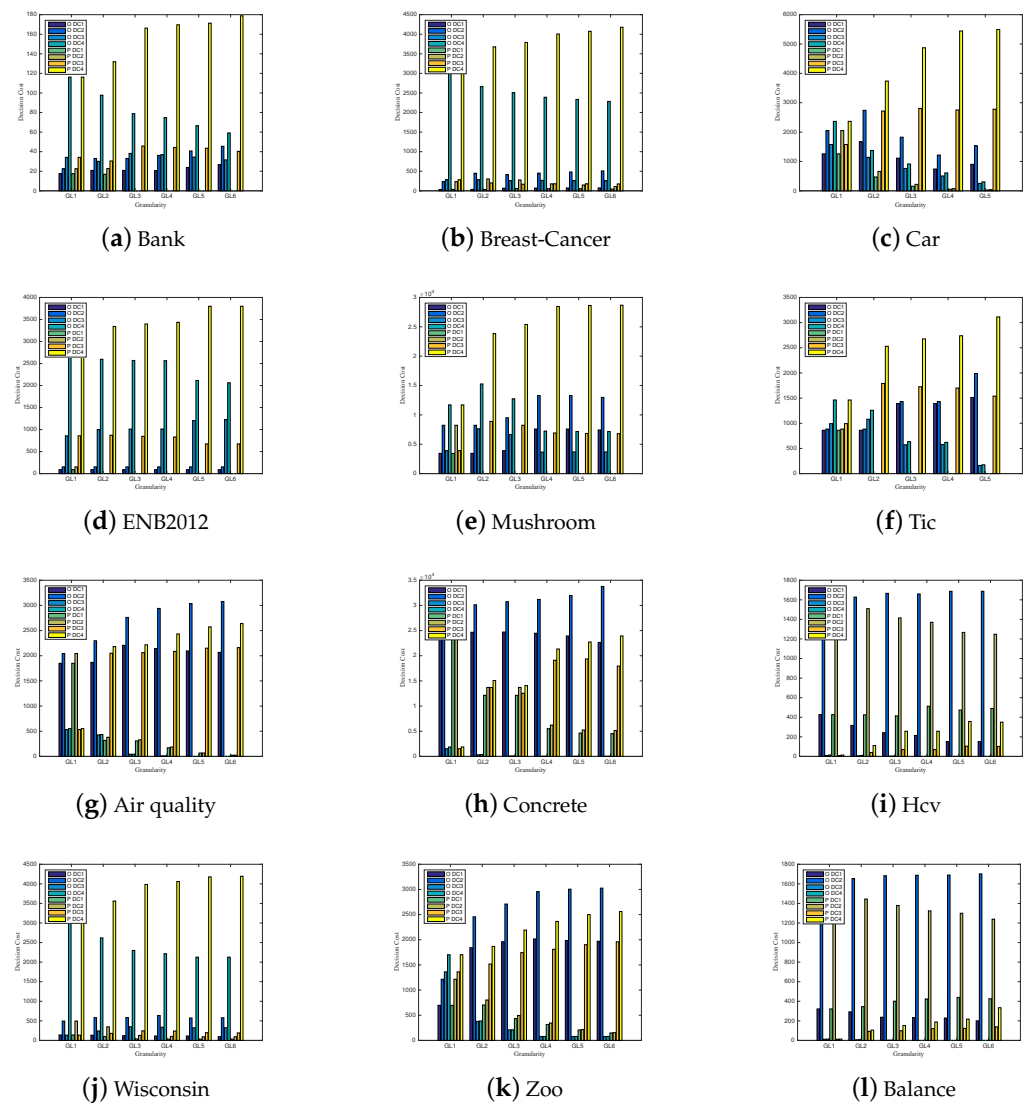**(j)** Wisconsin

(k) Zoo

(l) Balance

Figure 5. The misclassification cost of two boundary regions under different granularities.

In Figure 6, the horizontal and vertical axes denote the granularity and misclassification costs, respectively. $O\ DC\_lower$, $O\ DC\_upper$ and $O\ DC$ represent $DC(\underline{\sum_{i=1}^{m} A_i^O(X)})$, $DC(\overline{\sum_{i=1}^{m} A_i^O(X)})$ and $DC(\sum_{i=1}^{m} A_i^O(X))$. $P\ DC\_lower$, $P\ DC\_upper$ and $P\ DC$ represent $DC(\underline{\sum_{i=1}^{m} A_i^P(X)})$, $DC(\overline{\sum_{i=1}^{m} A_i^P(X)})$ and $DC(\sum_{i=1}^{m} A_i^P(X))$, respectively, namely, the misclassification costs generated when $\underline{\sum_{i=1}^{m} A_i^O(X)}$, $\overline{\sum_{i=1}^{m} R_i^O(X)}$ and $\sum_{i=1}^{m} A_i^O(X)$ are approximated to X and the misclassification costs generated when $\underline{\sum_{i=1}^{m} A_i^P(X)}$, $\overline{\sum_{i=1}^{m} A_i^P(X)}$ and $\sum_{i=1}^{m} A_i^P(X)$ are approximated to X. Obviously, compared with $\underline{\sum_{i=1}^{m} A_i^O(X)}$ and $\overline{\sum_{i=1}^{m} A_i^O(X)}$, the misclassification costs of $\sum_{i=1}^{m} A_i^O(X)$ are the least on each granularity. Similarly, compared

with $\sum\limits_{i=1}^{m} A_i^P(X), \overline{\sum\limits_{i=1}^{m} A_i^P(X)}$, the misclassification costs of $\sum\limits_{\underline{i=1}}^{m} A_i^P(X)$ are the least on each granularity. This is consistent with Theorems 5 and 6.
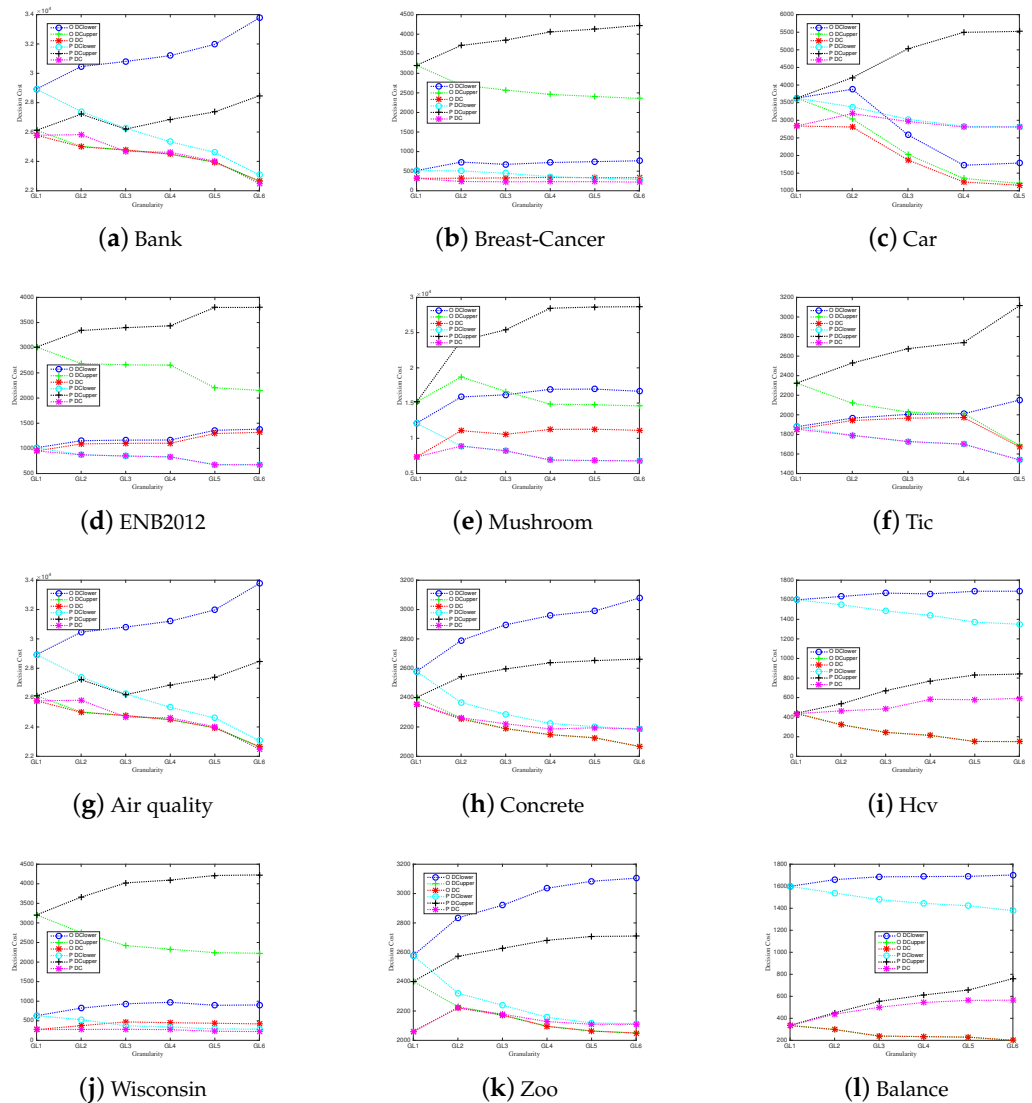

(a) Bank   **(b)** Breast-Cancer   **(c)** Car

(d) ENB2012   **(e)** Mushroom   **(f)** Tic

(g) Air quality   **(h)** Concrete   **(i)** Hcv

(j) Wisconsin   **(k)** Zoo   **(l)** Balance

Figure 6. The misclassification costs of $\sum\limits_{\underline{i=1}}^{m} A_i^O(X), \overline{\sum\limits_{i=1}^{m} A_i^O(X)}, \sum\limits_{i=1}^{m} A_i^O(X), \sum\limits_{i=1}^{m} A_i^P(X), \overline{\sum\limits_{i=1}^{m} A_i^P(X)}, \sum\limits_{i=1}^{m} A_i^P(X)$ with the changing granularity of each dataset.

6.2. Results and Discussions

According to the above experiments, compared with the upper/lower approximation sets, we can conclude that the multigranulation approximations have the following advantages when applied to decision-making environments:

(1) The misclassification costs of the approximation model monotonously decrease with the granularity being finer;

(2) In multigranulation approximations, under different granular layers, the misclassification costs incurred by the equivalence classes in approximating *X* are less than or equal to the misclassification costs incurred by the equivalence classes when they do not characterize *X* in boundary region I of optimistic and pessimistic rough sets. Moreover, the misclassification costs incurred by equivalence classes when they do not characterize *X* are less than or equal to the misclassification costs incurred by

the equivalence classes in approximating X in boundary region II of optimistic and pessimistic rough sets;

(3) Compared with the upper/lower approximation sets, the misclassification costs of the multigranulation approximations are the least on each granularity.

7. Conclusions

In MGRS, optimistic and pessimistic upper/lower approximation boundaries are utilized to characterize uncertain concepts. They still cannot take advantage of the known equivalence classes to establish the approximation of an uncertain concept. To handle the problem, cost-sensitive multigranulation approximations of rough sets were constructed. Furthermore, an optimization mechanism of the multigranulation approximations is proposed, which selects the optimal approximation to obtain the minimum misclassification costs under the conditions. The case study shows that the proposed algorithm is capable of searching for a rational approximation under restraints. Finally, the experiments demonstrate that the multigranulation approximations possess the least misclassification costs. In particular, our models apply to the decision-making environment where each decision-maker is independent. Moreover, our models are useful for extracting decision rules from distributive information systems and groups of intelligent agents through rough set approaches [34,36]. Figure 7 presents a diagram that summarizes the works conducted in this paper. Herein, we present the process of the cost-sensitive multigranulation approximations of rough sets; according to different granulation mechanisms, our approach can be extended to uncertainty models, i.e., vague sets, shadow sets, and neighborhood rough sets. These results will be important to contribute to the progress of the GrC theory.
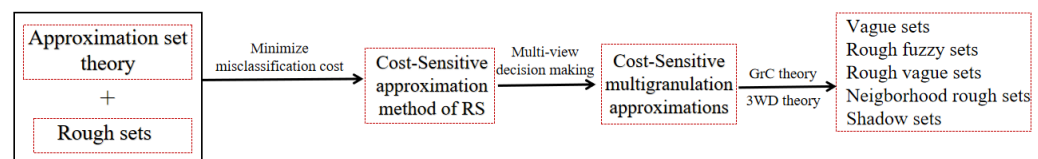

Figure 7. Diagram of works conducted in this paper.

Our future work will focus on the following two aspects: (1) We hope to build a more reasonable three-way decision model based on our model from optimistic and pessimistic perspectives; (2) we wish to combine the model with the cloud model theory to extend our model to construct a multigranulation approximation model with bidirectional cognitive computing. This will offer more cognitive advantages and benefits in application fields with uncertainty from multiple perspectives, i.e., image segmentation, clustering, and recommendation systems.

Author Contributions: Conceptualization, J.Y.; methodology, J.Y., J.K. and Q.L.; writing—original draft, J.Y. and J.K.; writing—review and editing, J.Y., J.K., Q.L. and Y.L.; data curation, J.Y., Q.L. and Y.L.; supervision, Y.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zadeh, L.A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst.* **1997**, *90*, 111–127. [CrossRef]
2. Bello, M.; Nápoles, G.; Vanhoof, K.; Bello, R. Data quality measures based on granular computing for multi-label classification. *Inf. Sci.* **2021**, *560*, 51–67. [CrossRef]
3. Pedrycz, W.; Chen, S. *Interpretable Artificial Intelligence: A Perspective of Granular Computing*; Springer Nature: Berlin/Heidelberg, Germany, 2021; Volume 937.
4. Li, J.; Mei, C.; Xu, W.; Qian, Y. Concept learning via granular computing: A cognitive viewpoint. *IEEE Trans. Fuzzy Syst.* **2015**, *298*, 447–467. [CrossRef] [PubMed]
5. Zadeh, L. Fuzzy sets. *Inf. Control.* **1965**, *8*, 338–353. [CrossRef]
6. Pawlak, Z. Rough sets. *Int. J. Comput. Inf. Sci.* **1982**, *11*, 341–356. [CrossRef]
7. Zhang, L.; Zhang, B. The quotient space theory of problem solving. In Proceedings of the International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, Chongqing, China, 26–29 May 2003; pp. 11–15.
8. Li, D.Y.; Meng, H.J.; Shi, X.M. Membership clouds and membership cloud generators. *J. Comput. Res. Dev.* **1995**, *32*, 15–20.
9. Colas-Marquez, R.; Mahfouf, M. Data Mining and Modelling of Charpy Impact Energy for Alloy Steels Using Fuzzy Rough Sets. *IFAC-Pap.* **2017**, *50*, 14970–14975. [CrossRef]
10. Hasegawa, K.; Koyama, M.; Arakawa, M.; Funatsu, K. Application of data mining to quantitative structure-activity relationship using rough set theory. *Chemom. Intell. Lab. Syst.* **2009**, *99*, 66–70. [CrossRef]
11. Santra, D.; Basu, S.K.; Mandal, J.K.; Goswami, S. Rough set based lattice structure for knowledge representation in medical expert systems: Low back pain management case study. *Expert Syst. Appl.* **2020**, *145*, 113084. [CrossRef]
12. Chebrolu, S.; Sanjeevi, S.G. Attribute Reduction in Decision-Theoretic Rough Set Model using Particle Swarm Optimization with the Threshold Parameters Determined using LMS Training Rule. *Procedia Comput. Sci.* **2015**, *57*, 527–536. [CrossRef]
13. Abdolrazzagh-Nezhad, M.; Radgohar, H.; Salimian, S.N. Enhanced cultural algorithm to solve multi-objective attribute reduction based on rough set theory. *Math. Comput. Simul.* **2020**, *170*, 332–350. [CrossRef]
14. Beaubier, S.; Defaix, C.; Albe-Slabi, S.; Aymes, A.; Galet, O.; Fournier, F.; Kapel, R. Multiobjective decision making strategy for selective albumin extraction from a rapeseed cold-pressed meal based on Rough Set approach. *Food Bioprod. Process.* **2022**, *133*, 34–44. [CrossRef]
15. Landowski, M.; Landowska, A. Usage of the rough set theory for generating decision rules of number of traffic vehicles. *Transp. Res. Procedia* **2019**, *39*, 260–269. [CrossRef]
16. Tawhid, M.; Ibrahim, A. Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 573–602. [CrossRef]
17. Zhang, Q.H.; Wang, G.Y.; Yu, X. Approximation sets of rough sets. *J. Softw.* **2012**, *23*, 1745–1759. [CrossRef]
18. Zhang, Q.H.; Wang, J.; Wang, G.Y. The approximate representation of rough-fuzzy sets. *Chin. J. Comput. Jisuanji Xuebao* **2015**, *38*, 1484–1496.
19. Zhang, Q.; Wang, J.; Wang, G.; Yu, H. The approximation set of a vague set in rough approximation space. *Inf. Sci.* **2015**, *300*, 1–19. [CrossRef]
20. Zhang, Q.H.; Zhang, P.; Wang, G.Y. Research on approximation set of rough set based on fuzzy similarity. *J. Intell. Fuzzy Syst.* **2017**, *32*, 2549–2562. [CrossRef]
21. Zhang, Q.H.; Yang, J.J.; Yao, L.Y. Attribute reduction based on rough approximation set in algebra and information views. *IEEE Access* **2016**, *4*, 5399–5407. [CrossRef]
22. Yao, L.Y.; Zhang, Q.H.; Hu, S.P.; Zhang, Q. Rough entropy for image segmentation based on approximation sets and particle swarm optimization. *J. Front. Comput. Sci. Technol.* **2016**, *10*, 699–708.
23. Zhang, Q.H.; Liu, K.X.; Gao, M. Approximation sets of rough sets and granularity optimization algorithm based on cost-sensitive. *J. Control. Decis.* **2020**, *35*, 2070–2080.
24. Yang, J.; Yuan, L.; Luo, T. Approximation set of rough fuzzy set based on misclassification cost. *J. Chongqing Univ. Posts Telecommun. (Nat. Sci. Ed.)* **2021**, *33*, 780–791.
25. Yang, J.; Luo, T.; Zeng, L.J.; Jin, X. The cost-sensitive approximation of neighborhood rough sets and granular layer selection. *J. Intell. Fuzzy Syst.* **2022**, *42*, 3993–4003. [CrossRef]
26. Siminski, K. 3WDNFS—Three-way decision neuro-fuzzy system for classification. *Fuzzy Sets Syst.* **2022**, *in press*. [CrossRef]
27. Subhashini, L.; Li, Y.; Zhang, J.; Atukorale, A.S. Assessing the effectiveness of a three-way decision-making framework with multiple features in simulating human judgement of opinion classification. *Inf. Process. Manag.* **2022**, *59*, 102823. [CrossRef]
28. Subhashini, L.; Li, Y.; Zhang, J.; Atukorale, A.S. Integration of semantic patterns and fuzzy concepts to reduce the boundary region in three-way decision-making. *Inf. Sci.* **2022**, *595*, 257–277. [CrossRef]

29. Mondal, A.; Roy, S.K.; Pamucar, D. Regret-based three-way decision making with possibility dominance and SPA theory in incomplete information system. *Expert Syst. Appl.* **2023**, *211*, 118688. [CrossRef]

30. Yao, Y.Y. Symbols-Meaning-Value (SMV) space as a basis for a conceptual model of data science. *Int. J. Approx. Reason.* **2022**, *144*, 113–128. [CrossRef]

31. Qian, Y.H.; Liang, J.Y.; Dang, C.Y. Incomplete multigranulation rough set. *IEEE Trans. Syst. Man-Cybern.-Part Syst. Humans* **2009**, *40*, 420–431. [CrossRef]

32. Huang, B.; Guo, C.X.; Zhuang, Y.L.; Li, H.X.; Zhou, X.Z. Intuitionistic fuzzy multigranulation rough sets. *Inf. Sci.* **2014**, *277*, 299–320. [CrossRef]

33. Li, F.J.; Qian, Y.H.; Wang, J.T.; Liang, J. Multigranulation information fusion: A Dempster-Shafer evidence theory-based clustering ensemble method. *Inf. Sci.* **2017**, *378*, 389–409. [CrossRef]

34. Liu, X.; Qian, Y.H.; Liang, J.Y. A rule-extraction framework under multigranulation rough sets. *Int. J. Mach. Learn. Cybern.* **2014**, *5*, 319–326. [CrossRef]

35. Liu, K.; Li, T.; Yang, X.; Ju, H.; Yang, X.; Liu, D. Hierarchical neighborhood entropy based multi-granularity attribute reduction with application to gene prioritization. *Int. J. Approx. Reason.* **2022**, *148*, 57–67. [CrossRef]

36. Qian, Y.H.; Liang, X.Y.; Lin, G.P.; Guo, Q.; Liang, J. Local multigranulation decision-theoretic rough sets. *Int. J. Approx. Reason.* **2017**, *82*, 119–137. [CrossRef]

37. Qian, Y.H.; Zhang, H.; Sang, Y.L.; Liang, J. Multigranulation decision-theoretic rough sets. *Int. J. Approx. Reason.* **2014**, *55*, 225–237. [CrossRef]

38. Xu, W.; Yuan, K.; Li, W. Dynamic updating approximations of local generalized multigranulation neighborhood rough set. *Appl. Intell.* **2022**, *52*, 9148–9173. [CrossRef]

39. Sun, L.; Wang, L.; Ding, W.; Qian, Y.; Xu, J. Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets. *IEEE Trans. Fuzzy Syst.* **2020**, *29*, 19–33. [CrossRef]

40. She, Y.H.; He, X.L.; Shi, H.X.; Qian, Y. A multiple-valued logic approach for multigranulation rough set model. *Int. J. Approx. Reason.* **2017**, *82*, 270–284. [CrossRef]

41. Li, W.; Xu, W.; Zhang, X.Y.; Zhang, J. Updating approximations with dynamic objects based on local multigranulation rough sets in ordered information systems. *Artif. Intell. Rev.* **2021**, *55*, 1821–1855. [CrossRef]

42. Zhang, C.; Li, D.; Zhai, Y.; Yang, Y. Multigranulation rough set model in hesitant fuzzy information systems and its application in person-job fit. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 717–729. [CrossRef]

43. Hu, C.; Zhang, L.; Wang, B.; Zhang, Z.; Li, F. Incremental updating knowledge in neighborhood multigranulation rough sets under dynamic granular structures. *Knowl.-Based Syst.* **2019**, *163*, 811–829. [CrossRef]

44. Hu, C.; Zhang, L. Dynamic dominance-based multigranulation rough sets approaches with evolving ordered data. *Int. J. Mach. Learn. Cybern.* **2021**, *12*, 17–38. [CrossRef]