

Neighborhood-Based Set-Valued Double-Quantitative Rough Sets

LI Wen-tao, LI Zhang, ZHU Chun-long, XU Wei-hua

(College of Artificial Intelligence, Southwest University, Chongqing 400715, China)

Abstract: Double-quantitative rough approximation, containing two types of quantitative information, indicated stronger generalization ability and more accurate data processing capacity than the single-quantitative rough approximation. In this paper, the neighborhood-based double-quantitative rough set models are firstly presented in a set-valued information system. Secondly, the attribute reduction method based on the lower approximation invariant is addressed, and the relevant algorithm for the approximation attribute reduction is provided in the set-valued information system. Finally, to illustrate the superiority and the effectiveness of the proposed reduction approach, experimental evaluation is performed using three datasets coming from the University of California-Irvine (UCI) repository.

Keywords: Attribute reduction; Double quantification; Fuzzy similarity relation; Set-valued information system

2000 MR Subject Classification: 03E72, 68T30, 68U35

CLC number: TP18 **Document code:** A

Article ID: 1002-0462 (2021) 02-0122-19

DOI: 10.13371/j.cnki.chin.q.j.m.2021.02.002

§1. Introduction

Rough set model [9] proposed by Pawlak, is a mathematical theory for data analysis. However, the traditional Pawlak rough set can only handle discrete data, and it is impossible to process the quantitative information. In order to make the rough set model suitable for

Received date: 2021-05-05

Foundation item: Supported by the College Students Innovation and Entrepreneurship Training Program project (Grant No. 101202010635586); National Natural Science Foundation of China (Grant No. 61772002, 61976245); Fundamental Research Funds for the Central Universities (Grant No. SWU119063); Scientific and Technological Project of Construction of Double City Economic Circle in Chengdu-Chongqing Area (Grant No. KJCX2020009); Science and Technology Research Program of Chongqing Education Commission (Grant No. KJQN202003806).

Biographies: LI Wen-tao (1987-), male, native of Xianning, Hubei, lecturer of Southwest University, engages in applied mathematics; LI Zhang (1999-), female, native of Jiaozuo, Henan, undergraduate student of Southwest University, engages in artificial intelligence; ZHU Chun-long (2000-), male, native of Ji'an, Jilin, undergraduate student of Southwest University, engages in artificial intelligence; XU Wei-hua (1979-), male, native of Hunyuan, Shanxi, professor of Southwest University, engages in applied mathematics.

more kinds of data processing scenarios, many scholars have extended the rough set model from multiple perspectives. Yao et al. proposed the rough set decision model based on Bayesian risk analysis [13, 14], Wong and Ziarko raised the 0.5-probability rough set model [10], Ziarko put forward a variable precision rough set model [21], Sun and Gong extended the variable precision probabilistic rough set model [11], Li and Xu considered two types of quantitative information on the basis of the probability rough set model and the degree rough set model, and proposed the the double-quantitative rough set model [5]. From the results reported in [5], it could find that the double-quantitative rough set model has strong fault tolerance. The relative and absolute quantitative information are two kinds of quantification mythologies encountered in certain applications [3, 12, 18–20].

Attribute reduction is an important issue of knowledge discovery [7, 8, 15–17]. As the information system often contains redundant attributes, it is necessary to obtain the necessary attributes of the information system to extract useful and important information, so as to improve the decision-making efficiency of the model [1, 2, 4, 6]. In this paper, we want to establish a double-quantitative rough set model in an information system with the set-valued data and discover the decision-making rules based on the proposed models, and then propose a lower approximation reduction algorithm for the model.

The paper is organized as follows. Related concepts are reviewed briefly in Section 2. In Section 3, we construct a fuzzy similarity relation in the set-valued information system, and derive the corresponding fuzzy similarity classes. And then present the neighborhood-based set-valued single-quantitative rough sets. In Section 4, we mainly discuss the neighborhood-based set-valued double-quantitative rough set models based on the fuzzy similarity relation. In Section 5, we develop the lower approximation reduction based on the presented double-quantitative rough set models, and provide the corresponding algorithm for the reduction method. In Section 6, we conduct a comprehensive analysis on the advantage and disadvantage for the single-quantitative and double-quantitative models. Finally, Section 7 covers some conclusions.

§2. Preliminary and basic concepts

This section briefly reviews the basic knowledge of Pawlak rough set, single-quantization models, and attribute reduction. In this paper, $P(U)$ and $F(U)$ represent the class of subsets and fuzzy subsets of U , respectively.

Definition 2.1. *Let (U, A, V, f) be an information system, R is an equivalence relation, for any $X \subseteq U$, the upper and lower approximation of the target set X are defined as*

$$\begin{aligned}\overline{R}(X) &= \{x \in U \mid [x]_R \cap X \neq \emptyset\}, \\ \underline{R}(X) &= \{x \in U \mid [x]_R \subseteq X\}.\end{aligned}$$

If $\overline{R}(X) = \underline{R}(X)$, X is a definable set; If $\overline{R}(X) \neq \underline{R}(X)$, X is a Pawlak rough set.

positive domain if $Neg_R(X) = \sim \overline{R}(X)$; the set $Bn_R(X)$ is called boundary domain if $Bn_R(X) = \overline{R}(X) - \underline{R}(X)$.

Definition 2.2. Let (U, A, V, f) be an information system, X is the target set, B is an equivalence relation, the upper and lower approximation operators based on the threshold α, β are defined as

$$\begin{aligned} \overline{B}_{(\alpha, \beta)}(X) &= \left\{ x \in U \mid \frac{|[x]_B \cap X|}{|[x]_B|} > \beta \right\}, \\ \underline{B}_{(\alpha, \beta)}(X) &= \left\{ x \in U \mid \frac{|[x]_B \cap X|}{|[x]_B|} \geq \alpha \right\}, \end{aligned}$$

where $0 \leq \beta < \alpha \leq 1$. If $\overline{B}_{(\alpha, \beta)}(X) \neq \underline{B}_{(\alpha, \beta)}(X)$, X is a probabilistic rough set (PRS) model.

Definition 2.3. Let (U, A, V, f) be an information system, for $B \subseteq A$, B is an equivalence relation, given a parameter k to indicate the degree, which satisfies $0 \leq k \leq |U|$, $k \in \mathbb{R}$, \mathbb{R} is a set of real numbers, for any $X \subseteq U$, the upper approximation operator and the lower approximation operator are defined as

$$\begin{aligned} \overline{B}_k(X) &= \{x \in U \mid |[x]_B \cap X| > k\}, \\ \underline{B}_k(X) &= \{x \in U \mid |[x]_B| - |[x]_B \cap X| \leq k\}. \end{aligned}$$

If $\overline{B}_k(X) = \underline{B}_k(X)$ then X is a degree k definable set; otherwise X is a degree k rough set, then there is no inclusion relation between the upper approximation and the lower approximation.

Definition 2.4. Let B be the coordination set of (U, A, F) for $B \subseteq A$, if $R_B = R_A$, and for any $b \in B$, exists $R_{B-\{b\}} \neq R_A$, then B is the reduction of (U, A, F) . For any information system (U, A, F) , its reduction always exists, and the intersection of all reductions is called the kernel of the information system.

An information system can have multiple reductions, but the kernel is unique.

§3. Neighborhood-based set-valued single-quantitative rough sets

In this section, we mainly study how to construct neighborhood-based set-valued single-quantitative rough set (NB-SV-SqRS) models. The first step in model construction is to construct the fuzzy similarity relation on the set-valued information system, and obtain the corresponding fuzzy similarity classes. As we know, the distance between two objects could reflect the degree of difference for this two objects. In a sense, the greater the distance between two objects, the lower similarity degree between the two objects. It is necessary to construct a distance function.

Definition 3.1. Let $S = (U, A, V, f)$ be a set-valued information system, x_i and x_j are two objects in U . If $B \subseteq A$, the distance d_{ij} of x_i, x_j about B is defined as:

$$d_{ij} = \left(\sum_{k=1}^{|B|} \left| \frac{|f(x_i, a_k) \oplus f(x_j, a_k)|}{|f(x_i, a_k) \cup f(x_j, a_k)|} \right|^p \right)^{1/p},$$

where \oplus is the symmetric difference operation of the set, if the value of p is different, the distance represented is different. Choose $p = 2$ at here. The distance function has the following properties:

- (1) *Non-negativity*: $d_{ij} \geq 0$,
- (2) *Symmetric*: $d_{ij} = d_{ji}$,
- (3) *Triangle inequality*: $d_{ij} \leq d_{im} + d_{mj}$.

It is clear that property (1), (2) was established, and now give the proof of property (3) by contradiction. Suppose $\forall x_i, x_m, x_j \subseteq U$, let $a = |f(x_i, a_k) \oplus f(x_m, a_k)|$, $b = |f(x_m, a_k) \oplus f(x_j, a_k)|$, $c = |f(x_i, a_k) \cup f(x_m, a_k)|$, $d = |f(x_m, a_k) \cup f(x_j, a_k)|$. Assume property 3 is not true, so

$$d_{ij} \geq d_{im} + d_{mj}. \tag{a}$$

Equation *a* can be defined as

$$\begin{aligned} \left| \frac{|f(x_i, a_k) \oplus f(x_j, a_k)|}{|f(x_i, a_k) \cup f(x_j, a_k)|} \right|^2 &\geq \left| \frac{a}{c} \right|^2 + \left| \frac{b}{d} \right|^2 + 2 \left(\frac{ab}{cd} \right)^{1/2}, \\ \left| \frac{|f(x_i, a_k) \oplus f(x_j, a_k)|}{|f(x_i, a_k) \cup f(x_j, a_k)|} \right|^2 &\geq \left| \frac{a}{c} \right|^2 + \left| \frac{b}{d} \right|^2 + 2 \left| \frac{ab}{cd} \right|. \end{aligned} \tag{b}$$

According to the knowledge of set theory can find that $A \oplus B \subseteq (A \oplus C) \cup (B \oplus C)$ and $A \cup B \subseteq (A \cup C) \cup (B \cup C)$, thus

$$\begin{aligned} |A \oplus B| &\leq |A \oplus C| + |B \oplus C|, \\ |A \cup B| &\leq |A \cup C| + |B \cup C|. \end{aligned}$$

Accordingly

$$d_{ij} \leq \left(\sum_{k=1}^{|B|} \left| \frac{a+b}{c+d} \right|^2 \right)^{1/2},$$

thus

$$\left| \frac{|f(x_i, a_k) \oplus f(x_j, a_k)|}{|f(x_i, a_k) \cup f(x_j, a_k)|} \right|^2 \leq \left| \frac{a}{c+d} \right|^2 + \left| \frac{b}{c+d} \right|^2 + 2 \left| \frac{ab}{(c+d)^2} \right|.$$

Obviously, it can be seen that

$$\left| \frac{a}{c+d} \right|^2 \leq \left| \frac{a}{c} \right|^2, \quad \left| \frac{b}{c+d} \right|^2 \leq \left| \frac{b}{d} \right|^2, \quad \left| \frac{ab}{(c+d)^2} \right| \leq \left| \frac{ab}{cd} \right|.$$

Therefore

$$\left| \frac{a}{c+d} \right|^2 + \left| \frac{b}{c+d} \right|^2 + 2 \left| \frac{ab}{(c+d)^2} \right| \leq \left| \frac{a}{c} \right|^2 + \left| \frac{b}{d} \right|^2 + 2 \left| \frac{ab}{cd} \right|.$$

So, equation *b* is not existed, it's also signify that equation *a* is not established. In conclusion, property 3 is established.

In order to get the similar value s_{ij} between the two object x_i and x_j under the attribute set B through the distance function, however the value of s_{ij} should be between 0 and 1, so d_{ij} need to normalized.

Definition 3.2. Let (U, A, V, f) be a set-valued information system and $B \subseteq A$, d_{ij} is the distance between two objects x_i and x_j under the attribute set B , D_{ij} is the distance between the two objects x_i and x_j under the attribute set A , then normalize d_{ij} to d^*_{ij} :

$$d^*_{ij} = \frac{d_{ij}}{\max(D_{ij})}.$$

Definition 3.3. Let (U, A, V, f) be a set-valued information system and $B \subseteq A$, \tilde{B} is the neighborhood-based fuzzy similarity relationship of B , and its fuzzy similarity matrix $S(\tilde{B})$ is defined as:

$$S(\tilde{B}) = \begin{pmatrix} S_{11} & \dots & S_{1n} \\ \vdots & \ddots & \vdots \\ S_{n1} & \dots & S_{nn} \end{pmatrix},$$

where $s_{ij} = 1 - d^*_{ij}$ represents the similarity value between the two objects x_i and x_j , and the relationship \tilde{B} is the fuzzy similarity relationship.

Suppose I is a identity matrix, $S(\tilde{B})$ satisfies the following two properties:

- (1) Reflexivity: $I \leq S(\tilde{B}) (\Leftrightarrow s_{ii} = 1)$.
- (2) Symmetry: $S(\tilde{B})^T = S(\tilde{B}) (\Leftrightarrow s_{ij} = s_{ji})$.

Thus $S(\tilde{B})$ is a fuzzy similarity matrix, \tilde{B} is proved by the fuzzy similarity relationship.

Definition 3.4. Let (U, A, V, f) be a set-valued information system and $B \subseteq A$, \tilde{B} is the fuzzy similarity relation of B , for any $\forall x \in B$, $[x]_{\tilde{B}} = \{y | y \in B \wedge x \tilde{B} y\}$, $[x]_{\tilde{B}}$ is called the fuzzy similarity class of x_i about the fuzzy similarity relation \tilde{B} which is referred to as the fuzzy similarity class for short, abbreviated as $[x]$.

If \tilde{B} is the fuzzy similarity relation on set-valued information system (U, A, V, f) , then $[x]$ has the following properties:

- (1) $\forall x \in A$, $[x]$ is a nonempty subset.
- (2) $\cup\{[x] | x \in A\} = A$.

The previous section explained how to construct a fuzzy similarity relation. The following will discuss how to construct two single-quantitative rough set models. The difference between the two models is that the introduced quantitative information is different. The introduction of relative quantitative information can construct a neighborhood-based set-valued probabilistic single-quantitative rough set (NB-SVP-SqRS); the introduction of absolute quantitative information can construct a neighborhood-based set-valued degree single-quantitative rough set (NB-SVD-SqRS).

Definition 3.5. Let (U, A, V, f) be a set-valued information system and $B \subseteq A$, \tilde{B} is the fuzzy similarity, for any $\tilde{X} \in \mathcal{F}(U)$, $\mathcal{F}(U)$ represents the class of all fuzzy subsets of U , let $0 \leq \beta < \alpha \leq 1$, then the upper approximation operator and lower approximation operator of the probabilistic set-valued information system are

$$\overline{\tilde{B}}_{(\alpha, \beta)}(\tilde{X}) = \left\{ x \in U \mid \frac{|[x]_{\tilde{B}} \cap \tilde{X}|}{|[x]_{\tilde{B}}|} > \beta \right\},$$

$$\underline{\tilde{B}}_{(\alpha,\beta)}(\tilde{X}) = \left\{ x \in U \mid \frac{|[x]_{\tilde{B}} \cap \tilde{X}|}{|[x]_{\tilde{B}}|} \geq \alpha \right\}.$$

The upper approximation operator and lower approximation operator of the probabilistic set-valued information system can also be defined as

$$\begin{aligned} \overline{\tilde{B}}_{(\alpha,\beta)}(\tilde{X}) &= \left\{ x \in U \mid |[x]_{\tilde{B}} \cap \tilde{X}| > \beta \cdot |[x]_{\tilde{B}}| \right\}, \\ \underline{\tilde{B}}_{(\alpha,\beta)}(\tilde{X}) &= \left\{ x \in U \mid |[x]_{\tilde{B}} \cap \tilde{X}| \geq \alpha \cdot |[x]_{\tilde{B}}| \right\}. \end{aligned}$$

It can be seen that there is a strict containment relationship between $\overline{\tilde{B}}_{(\alpha,\beta)}(\tilde{X})$ and $\underline{\tilde{B}}_{(\alpha,\beta)}(\tilde{X})$, which is $\underline{\tilde{B}}_{(\alpha,\beta)}(\tilde{X}) \subseteq \overline{\tilde{B}}_{(\alpha,\beta)}(\tilde{X})$.

Definition 3.6. Let (U, A, V, f) be a set-valued information system and $B \subseteq A$, \tilde{B} is the fuzzy similarity relation of B , $0 \leq k \leq |U|$, $k \in \mathbb{R}$, for any $\tilde{X} \in \mathcal{F}(U)$, then the upper approximation operator and lower approximation operator of the degree set-valued information system be defined as

$$\begin{aligned} \overline{\tilde{B}}_k(\tilde{X}) &= \left\{ x \in U \mid |[x]_{\tilde{B}} \cap \tilde{X}| > k \right\}, \\ \underline{\tilde{B}}_k(\tilde{X}) &= \left\{ x \in U \mid |[x]_{\tilde{B}}| - |[x]_{\tilde{B}} \cap \tilde{X}| \leq k \right\}. \end{aligned}$$

The lower operators of degree set-valued information system can also be defined as:

$$\underline{\tilde{B}}_k(\tilde{X}) = \left\{ x \in U \mid |[x]_{\tilde{B}} \cap \tilde{X}| \geq |[x]_{\tilde{B}}| - k \right\}.$$

Obviously, compared with the classic rough set model, there is not strict inclusion relation between $\overline{\tilde{B}}_k(\tilde{X})$ and $\underline{\tilde{B}}_k(\tilde{X})$. If $\overline{\tilde{B}}_k(\tilde{X}) = \underline{\tilde{B}}_k(\tilde{X})$, then \tilde{X} is a definable set, otherwise, \tilde{X} is the degree k rough set.

§4. Neighborhood-based set-valued double-quantitative rough sets

Let (U, A, V, f) be a set-valued information system and $B \subseteq A$, \tilde{B} is the fuzzy similarity relation of B . $0 \leq \beta < \alpha \leq 1$ and $0 \leq k \leq |U|$, $k \in \mathbb{R}$. For any $\tilde{X} \in \mathcal{F}(U)$, let

$$P = \frac{|[x]_{\tilde{B}} \cap \tilde{X}|}{|[x]_{\tilde{B}}|}, \quad D = |[x]_{\tilde{B}} \cap \tilde{X}|.$$

The logical operators are introduced into the upper and lower approximations, and four neighborhood-based double-quantitative rough approximations are obtained. There are two kinds of upper approximation operators, which can be shown as follows.

(1) The upper approximation operator with logical “and” $\overline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X}) = \overline{\tilde{B}}_{(\alpha,\beta)}(\tilde{X}) \cap \underline{\tilde{B}}_k(\tilde{X})$. The upper approximation operator with logical “and” also can be defined as

$$\begin{aligned} \overline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X}) &= \{x \in U \mid P > \beta \wedge D > k\}, \\ \overline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X}) &= \{x \in U \mid D > \max(\beta \cdot |[x]_{\tilde{B}}|, k)\}. \end{aligned}$$

(2) The upper approximation operator with logical “or”. $\overline{\tilde{B}_{(\alpha,\beta,k)}^\vee}(\tilde{X}) = \overline{\tilde{B}_{(\alpha,\beta)}}(\tilde{X}) \cup \overline{\tilde{B}_k}(\tilde{X})$.
 The upper approximation operator with logical “or” also can be defined as

$$\begin{aligned} \overline{\tilde{B}_{(\alpha,\beta,k)}^\vee}(\tilde{X}) &= \{x \in U \mid P > \beta \vee G > k\}, \\ \overline{\tilde{B}_{(\alpha,\beta,k)}^\vee}(\tilde{X}) &= \{x \in U \mid D > \min(\beta \cdot |[x]_{\tilde{B}}|, k)\}. \end{aligned}$$

There are two lower approximation operators, which are:

(1) The lower approximation operator with logical “and”, $\underline{\tilde{B}_{(\alpha,\beta,k)}^\wedge}(\tilde{X}) = \underline{\tilde{B}_{(\alpha,\beta)}}(\tilde{X}) \cap \underline{\tilde{B}_k}(\tilde{X})$.
 The lower approximation operator with logical “and” can also be defined as,

$$\begin{aligned} \underline{\tilde{B}_{(\alpha,\beta,k)}^\wedge}(\tilde{X}) &= \{x \in U \mid P \geq \alpha \wedge |[x]_{\tilde{B}}| - D \leq k\}, \\ \underline{\tilde{B}_{(\alpha,\beta,k)}^\wedge}(\tilde{X}) &= \{x \in U \mid D > \max(\alpha \cdot |[x]_{\tilde{B}}|, |[x]_{\tilde{B}}| - k)\}. \end{aligned}$$

(2) The lower approximation operator with logical “or”, $\underline{\tilde{B}_{(\alpha,\beta,k)}^\vee}(\tilde{X}) = \underline{\tilde{B}_{(\alpha,\beta)}}(\tilde{X}) \cup \underline{\tilde{B}_k}(\tilde{X})$.
 The lower approximation operator with logical “or” can also be defined as

$$\begin{aligned} \underline{\tilde{B}_{(\alpha,\beta,k)}^\vee}(\tilde{X}) &= \{x \in U \mid P \geq \alpha \vee |[x]_{\tilde{B}}| - D \leq k\}, \\ \underline{\tilde{B}_{(\alpha,\beta,k)}^\vee}(\tilde{X}) &= \{x \in U \mid D > \min(\alpha \cdot |[x]_{\tilde{B}}|, |[x]_{\tilde{B}}| - k)\}. \end{aligned}$$

Based on the above four approximation operators, optimistic and pessimistic double-quantitative rough set models are proposed here, the upper and lower approximations of neighborhood-based set-valued pessimistic double-quantitative rough sets model (NB-SVP-DqRS) both choose the approximation operators with logical “and”, denoted by $(U, \overline{\tilde{B}_{(\alpha,\beta,k)}^\wedge}(\tilde{X}), \underline{\tilde{B}_{(\alpha,\beta,k)}^\wedge}(\tilde{X}))$; the upper and lower approximations of neighborhood-based set-valued optimistic double-quantitative rough sets model (NB-SVO-DqRS) both choose the approximation operators with logical “or”, denoted by $(U, \overline{\tilde{B}_{(\alpha,\beta,k)}^\vee}(\tilde{X}), \underline{\tilde{B}_{(\alpha,\beta,k)}^\vee}(\tilde{X}))$. Due to the introduction of different logical operators in the models, compared with the NB-SVO-DqRS model, the upper and lower approximation conditions of the NB-SVP-DqRS model are more stringent, therefore the classification conditions are also more stringent. The two models will be elaborated as follows.

Definition 4.1. Let (U, A, V, f) be a set-valued information system and $B \subseteq A$, \tilde{B} is the fuzzy similarity relation, for any $\tilde{X} \in \mathcal{F}(U)$, $x \in U$, the NB-SVP-DqRS model is defined as

$$\begin{aligned} \overline{\tilde{B}_{(\alpha,\beta,k)}^\wedge}(\tilde{X}) &= \{x \in U \mid P > \beta \wedge D > k\}, \\ \underline{\tilde{B}_{(\alpha,\beta,k)}^\wedge}(\tilde{X}) &= \{x \in U \mid P \geq \alpha \wedge |[x]_{\tilde{B}}| - D \leq k\}. \end{aligned}$$

The NB-SVP-DqRS model can also be defined as

$$\begin{aligned} \underline{\tilde{B}_{(\alpha,\beta,k)}^\wedge}(\tilde{X}) &= \{x \in U \mid D > \max(\beta \cdot |[x]_{\tilde{B}}|, k)\}, \\ \underline{\tilde{B}_{(\alpha,\beta,k)}^\wedge}(\tilde{X}) &= \{x \in U \mid D > \max(\alpha \cdot |[x]_{\tilde{B}}|, |[x]_{\tilde{B}}| - k)\}. \end{aligned}$$

Definition 4.2. Let (U, B, V, f) be a set-valued information, and \tilde{B} is the fuzzy similarity relation, the NB-SVP-DqRS model can be represented as $(U, \overline{\tilde{B}_{(\alpha, \beta, k)}^\wedge}(\tilde{X}), \underline{\tilde{B}_{(\alpha, \beta, k)}^\wedge}(\tilde{X}))$, Obviously there is no strict inclusion relationship between the upper and lower approximations of this model, so the positive domain, negative domain, upper boundary domain, lower boundary domain of NB-SVP-DqRS model are defined as:

$$\begin{aligned} Pos_{(\alpha, \beta, k)}^P(\tilde{X}) &= \overline{\tilde{B}_{(\alpha, \beta, k)}^\wedge}(\tilde{X}) \cap \underline{\tilde{B}_{(\alpha, \beta, k)}^\wedge}(\tilde{X}), \\ Neg_{(\alpha, \beta, k)}^P(\tilde{X}) &= \sim(\overline{\tilde{B}_{(\alpha, \beta, k)}^\wedge}(\tilde{X}) \cap \underline{\tilde{B}_{(\alpha, \beta, k)}^\wedge}(\tilde{X})), \\ Ubn_{(\alpha, \beta, k)}^P(\tilde{X}) &= \overline{\tilde{B}_{(\alpha, \beta, k)}^\wedge}(\tilde{X}) - \underline{\tilde{B}_{(\alpha, \beta, k)}^\wedge}(\tilde{X}), \\ Lbn_{(\alpha, \beta, k)}^P(\tilde{X}) &= \underline{\tilde{B}_{(\alpha, \beta, k)}^\wedge}(\tilde{X}) - \overline{\tilde{B}_{(\alpha, \beta, k)}^\wedge}(\tilde{X}). \end{aligned}$$

Definition 4.3. Let (U, A, V, f) be a set-valued information system and $B \subseteq A$. \tilde{B} is the fuzzy similarity relation of B , for any $\tilde{X} \in \mathcal{F}(U)$, $x \in U$, the NB-SVO-DqRS model is defined as

$$\begin{aligned} \overline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}) &= \{x \in U \mid P > \beta \vee D > k\}, \\ \underline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}) &= \{x \in U \mid P \geq \alpha \vee |[x]_{\tilde{B}}| - D \leq k\}. \end{aligned}$$

The NB-SVO-DqRS model can also be defined as

$$\begin{aligned} \overline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}) &= \{x \in U \mid D > \min(\beta \cdot |[x]_{\tilde{B}}|, k)\}, \\ \underline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}) &= \{x \in U \mid D > \min(\alpha \cdot |[x]_{\tilde{B}}|, |[x]_{\tilde{B}}| - k)\}. \end{aligned}$$

Definition 4.4. Let (U, B, V, f) be a set-valued information system, \tilde{B} is the fuzzy similarity relation of attribute set B . Use $(U, \overline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}), \underline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}))$ to represent the NB-SVO-DqRS model. The positive domain, negative domain, upper boundary domain, lower boundary domain of NB-SVO-DqRS model are defined as

$$\begin{aligned} Pos_{(\alpha, \beta, k)}^O(\tilde{X}) &= \overline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}) \cap \underline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}), \\ Neg_{(\alpha, \beta, k)}^O(\tilde{X}) &= \sim(\overline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}) \cap \underline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X})), \\ Ubn_{(\alpha, \beta, k)}^O(\tilde{X}) &= \overline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}) - \underline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}), \\ Lbn_{(\alpha, \beta, k)}^O(\tilde{X}) &= \underline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}) - \overline{\tilde{B}_{(\alpha, \beta, k)}^\vee}(\tilde{X}). \end{aligned}$$

In the above two models, if the objects belonging to the positive region, we make the decision to accept; if the objects belonging to the negative region, we make the decision to reject; if the objects belonging to the boundary region indicate that we need to make further judgments before we can make a decision. Among them, the objects in the upper boundary domain are more inclined to reject, and the objects in the lower boundary domain are more inclined to accept.

In the following, we will analyze the two NB-SV-DqRS models with practical examples to obtain the upper and lower approximations and to discuss the decision rules of two models respectively.

Example 4.1. Given a set-valued information system $I=(U,A,V,F)$ (Table 1). The universe U is a universe of 6 products, five attributes a_1, a_2, a_3, a_4, a_5 , respectively represent the size, feel, weight, color, and shape of the product. Given a target set $\tilde{X} = \left\{ \frac{0.30}{x_1} + \frac{0.50}{x_2} + \frac{0.10}{x_3} + \frac{0.50}{x_4} + \frac{0.00}{x_5} + \frac{0.20}{x_6} \right\}$ represent each parameter of the product, set the parameter $\alpha=0.50, \beta=0.45, k=1.0$. Then we can discuss how to make decisions based on the NB-SV-DqRS models.

Table 1 Set-valued information system

U	a_1	a_2	a_3	a_4	a_5
x_1	{2, 4}	{1, 2}	{2, 3}	{1, 3, 4}	{2, 4, 5}
x_2	{2, 4}	{2}	{1}	{2, 4}	{2, 3}
x_3	{2}	{1}	{3}	{3, 4}	{1, 4, 5}
x_4	{2, 3}	{1, 2}	{1, 3}	{2, 4}	{1}
x_5	{2, 3, 5}	{1, 2, 3, 4}	{1, 2, 3}	{1, 2, 4, 5}	{2}
x_6	{2, 4, 5}	{2}	{1, 4, 5}	{2, 4, 5}	{1, 2, 4}

Calculate the distance between objects according to the distance function

$$D(A) = \begin{pmatrix} 0.00 & 1.54 & 1.05 & 1.57 & 1.31 & 1.50 \\ 1.54 & 0.00 & 1.92 & 1.39 & 1.44 & 1.11 \\ 1.05 & 1.92 & 0.00 & 1.28 & 1.76 & 1.80 \\ 1.57 & 1.39 & 1.28 & 0.00 & 1.31 & 1.39 \\ 1.31 & 1.44 & 1.76 & 1.31 & 0.00 & 1.40 \\ 1.50 & 1.11 & 1.80 & 1.39 & 1.40 & 0.00 \end{pmatrix}.$$

It is easy to find that $\max(D_{ij})=1.92$, then normalize the above matrix to

$$D^*(A) = \begin{pmatrix} 0.00 & 0.80 & 0.55 & 0.82 & 0.68 & 0.78 \\ 0.80 & 0.00 & 1.00 & 0.73 & 0.75 & 0.58 \\ 0.55 & 1.00 & 0.00 & 0.67 & 0.92 & 0.94 \\ 0.82 & 0.73 & 0.67 & 0.00 & 0.68 & 0.72 \\ 0.68 & 0.75 & 0.92 & 0.68 & 0.00 & 0.73 \\ 0.78 & 0.58 & 0.94 & 0.72 & 0.73 & 0.00 \end{pmatrix}.$$

The fuzzy similarity matrix is

$$S(\tilde{B}) = \begin{pmatrix} 1.00 & 0.20 & 0.45 & 0.18 & 0.32 & 0.22 \\ 0.20 & 1.00 & 0.00 & 0.27 & 0.25 & 0.42 \\ 0.45 & 0.00 & 1.00 & 0.33 & 0.08 & 0.06 \\ 0.18 & 0.27 & 0.33 & 1.00 & 0.32 & 0.28 \\ 0.32 & 0.25 & 0.08 & 0.32 & 1.00 & 0.27 \\ 0.22 & 0.42 & 0.06 & 0.28 & 0.27 & 1.00 \end{pmatrix}.$$

Then the fuzzy similarity classes of attribute set A could be obtained as follows.

$$\begin{aligned}
 [x_1]_{\tilde{B}} &= \left\{ \frac{1}{x_1} + \frac{0.20}{x_2} + \frac{0.45}{x_3} + \frac{0.18}{x_4} + \frac{0.32}{x_5} + \frac{0.22}{x_6} \right\} \\
 [x_2]_{\tilde{A}} &= \left\{ \frac{0.20}{x_1} + \frac{1}{x_2} + \frac{0}{x_3} + \frac{0.27}{x_4} + \frac{0.25}{x_5} + \frac{0.42}{x_6} \right\} \\
 [x_3]_{\tilde{A}} &= \left\{ \frac{0.45}{x_1} + \frac{0}{x_2} + \frac{1.00}{x_3} + \frac{0.33}{x_4} + \frac{0.08}{x_5} + \frac{0.06}{x_6} \right\} \\
 [x_4]_{\tilde{A}} &= \left\{ \frac{0.18}{x_1} + \frac{0.27}{x_2} + \frac{0.33}{x_3} + \frac{1.00}{x_4} + \frac{0.32}{x_5} + \frac{0.28}{x_6} \right\} \\
 [x_5]_{\tilde{A}} &= \left\{ \frac{0.32}{x_1} + \frac{0.25}{x_2} + \frac{0.08}{x_3} + \frac{0.32}{x_4} + \frac{1}{x_5} + \frac{0.27}{x_6} \right\} \\
 [x_6]_{\tilde{A}} &= \left\{ \frac{0.22}{x_1} + \frac{0.42}{x_2} + \frac{0.06}{x_3} + \frac{0.28}{x_4} + \frac{0.27}{x_5} + \frac{1.00}{x_6} \right\}.
 \end{aligned}$$

Consequently,

$$\begin{aligned}
 |[x_1]_{\tilde{R}}| &= 2.37, & |[x_2]_{\tilde{R}}| &= 2.14, & |[x_3]_{\tilde{R}}| &= 1.93, \\
 |[x_4]_{\tilde{R}}| &= 2.38, & |[x_5]_{\tilde{R}}| &= 2.24, & |[x_6]_{\tilde{R}}| &= 2.25.
 \end{aligned}$$

The value of $|[x_i]_{\tilde{B}} \cap \tilde{X}|$ are shown as

$$\begin{aligned}
 |[x_1]_{\tilde{B}} \cap \tilde{X}| &= 0.98, & |[x_2]_{\tilde{B}} \cap \tilde{X}| &= 0.97, & |[x_3]_{\tilde{B}} \cap \tilde{X}| &= 0.79, \\
 |[x_4]_{\tilde{B}} \cap \tilde{X}| &= 1.26, & |[x_5]_{\tilde{B}} \cap \tilde{X}| &= 1.15, & |[x_6]_{\tilde{B}} \cap \tilde{X}| &= 1.05.
 \end{aligned}$$

And the conditional probabilities of all objects are

$$\begin{aligned}
 \frac{|[x_1]_{\tilde{B}} \cap \tilde{X}|}{|[x_1]_{\tilde{B}}|} &= 0.415, & \frac{|[x_2]_{\tilde{B}} \cap \tilde{X}|}{|[x_2]_{\tilde{B}}|} &= 0.453, & \frac{|[x_3]_{\tilde{B}} \cap \tilde{X}|}{|[x_3]_{\tilde{B}}|} &= 0.411, \\
 \frac{|[x_4]_{\tilde{B}} \cap \tilde{X}|}{|[x_4]_{\tilde{B}}|} &= 0.528, & \frac{|[x_5]_{\tilde{B}} \cap \tilde{X}|}{|[x_5]_{\tilde{B}}|} &= 0.515, & \frac{|[x_6]_{\tilde{B}} \cap \tilde{X}|}{|[x_6]_{\tilde{B}}|} &= 0.469.
 \end{aligned}$$

Then the upper and lower approximations of the NB-SVP-DqRS model are expressed as

$$\begin{aligned}
 \overline{\tilde{B}}_{(\alpha, \beta, k)}^{\wedge}(\tilde{X}) &= \{x_4, x_5, x_6\}, \\
 \underline{\tilde{B}}_{(\alpha, \beta, k)}^{\wedge}(\tilde{X}) &= \{x_4, x_5\}.
 \end{aligned}$$

The corresponding positive region, negative region, upper boundary region, and lower boundary region are derived as follows,

$$\begin{aligned}
 Pos^P_{(\alpha, \beta, k)}(\tilde{X}) &= \{x_4, x_5\}, \\
 Neg^P_{(\alpha, \beta, k)}(\tilde{X}) &= \{x_1, x_2, x_3, x_6\}, \\
 Ubn^P_{(\alpha, \beta, k)}(\tilde{X}) &= \{x_6\}, \\
 Lbn^P_{(\alpha, \beta, k)}(\tilde{X}) &= \emptyset.
 \end{aligned}$$

The objects x_4, x_5 belonging to the positive region indicates that they are qualified; the objects x_1, x_2, x_3, x_6 belonging to the negative region indicates that their quality are unqualified; and the product x_6 belonging to the upper boundary region needs further inspection to determine whether it is qualified or unqualified, but it is more likely to be qualified; and here products belonging to the lower boundary domain don't exist.

And the upper and lower approximations of the NB-SVO-DqRS model are expressed as

$$\overline{B}_{(\alpha, \beta, k)}^{\vee}(\tilde{X}) = \{x_2, x_3, x_4, x_5, x_6\}, \quad \underline{B}_{(\alpha, \beta, k)}^{\vee}(\tilde{X}) = \{x_4, x_5\}.$$

Its positive region, negative region, upper boundary region, and lower boundary region are as follows,

$$\begin{aligned} Pos^O_{(\alpha, \beta, k)}(\tilde{X}) &= \{x_4, x_5\}, \\ Neg^O_{(\alpha, \beta, k)}(\tilde{X}) &= \{x_1, x_2, x_3, x_6\}, \\ Ubn^O_{(\alpha, \beta, k)}(\tilde{X}) &= \{x_2, x_3, x_6\}, \\ Lbn^O_{(\alpha, \beta, k)}(\tilde{X}) &= \emptyset. \end{aligned}$$

The products x_4, x_5 belonging to the positive region indicates that they are qualified; the products x_1, x_2, x_3, x_6 belonging to the negative region indicates that they are unqualified; and the products x_2, x_3, x_6 belonging to the upper boundary region need further inspection to determine whether it is qualified or unqualified, but they are more likely to be qualified; at here no product belongs to the lower boundary region.

§5. Lower approximation reduction based on double-quantitative rough set

In the previous section, two double-quantitative rough set models under the set-value information system were constructed, and decision rules were obtained. In this section, the lower approximation reduction methods of the two models will be discussed.

Definition 5.1. Let (U, A, V, f) be a set-valued information system, R is a fuzzy similarity relationship based on this information system. Where $B \subseteq A$ is an approximate reduction in a set-valued information system, the following two conditions must be established:

- (1) $\underline{R}_B(X) = \underline{R}_A(X)$,
- (2) For any $b \in B$, exist $\underline{R}_{B-\{b\}}(X) \neq \underline{R}_A(X)$.

Definition 5.2. For a set-valued information system (U, A, V, f) , its lower approximation reduction always exists.

For any $a \in A$, if both exist $\underline{R}_{A-\{a\}}(X) \neq \underline{R}_A(X)$, then A is a reduction. If $a \in A$, let $\underline{R}_{A-\{a\}}(X) = \underline{R}_A(X)$, then research $B_1 = A - \{a\}$, for any $b_1 \in B_1$, there is $\underline{R}_{B_1-\{b_1\}}(X) \neq \underline{R}_A(X)$, so B_1 is a reduction, if exist $b_1 \in B_1$, let $\underline{R}_{B_1-\{b_1\}}(X) = \underline{R}_A(X)$, then research $B_2 = B_1 - \{b_1\}$, and as so on. Since A is a finite set, the set B can always be found, let $\underline{R}_B(X) = \underline{R}_A(X)$, at this time, B is a reduction of the information system.

Recalling the previous section, two rough set models have been constructed in the set-valued information system, containing two lower approximation operators, namely $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ and $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\vee}(\tilde{X})$. The reductions of two models are all keep the lower approximation constant. In the following part, we just take the lower approximation reduction algorithm of the NB-SVP-DqRS model as an example to illustrate the two lower approximation reduction algorithms.

Algorithm 1: Lower approximation reduction based on invariance of $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\vee}(\tilde{X})$

Input : A set-valued decision information system $SDS (U, A \cup D, V, f)$.

Output : The reduction A .

- 1 **Step 1:** Construct fuzzy similar relation \tilde{B} and calculate the $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ of SDS .
 - 2 **Step 2:** A is the conditional attribute set, and only delete the attribute a_i ($a_i \in A$) every time on A , calculate the $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ of $(U, A - \{a_i\} \cup D, V, f)$, and compare with step 1.
 - 3 **Step 3:** If $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ changes, add a_i to the array B ; if $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ doesn't change, add a_i to the array K , K is kernel element. Let $i = i + 1$, determine whether there is an attribute a_i in the set A that has not been deleted. If true, go to step 3; if false, go to step 5.
 - 4 **Step 4:** Follow a sequence to add the attributes a_{B_i} which stored in the set B to the kernel element set K , calculate the $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ of $(U, K + \{a_{B_i}\} \cup D, V, f)$, and compare with step 1, if the result is the same, stop adding a_{B_i} to K , if the result is different, adding a_{B_i} to K .
 - 5 **Step 5:** Return A .
-

The kernel idea of this algorithm is to obtain the non-kernel element set B at first, and then sequentially add the attributes a_{B_i} which is stored in the set B to the kernel attribute set K , $a_{B_i} \in A - K$, until the $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ of $(U, A \cup D, V, f)$ is same as the $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ of $(U, A \cup D, V, f)$. At this time, $K + \{a_{B_i}\}$ is the reduction. The complexity analysis of algorithm 1 is conducted as follows.

Input a set-valued information system with $m \times n$ dimensions, the time complexity of algorithm step 1 is $O(m^2n^2)$. The time complexity of calculating fuzzy similarity relation \tilde{B} is $O(mn^2)$, the time complexity of calculating $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ is $O(m)$, the purpose of algorithm steps 2 and 3 is to find the kernel elements, their time complexity are $O(m)$, step 4 is to find the reduction, the best situation is that the set B is an empty set, and the kernel itself is a reduction, at this time the total time complexity is $O(m^2n^2)$; the worst situation is that algorithm needs to judge $m - i$ times, at this time, the process of calculating $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ needs to be looped $m - i$ times, so at this time the total time complexity is $O(m^3n^2)$.

The calculation idea of space complexity is similar to time complexity. In Step 1, only one array is needed in the construction of fuzzy similarity relationship \tilde{B} and the calculation of $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$, their space complexity are both $O(1)$. The best situation is that the set B is an empty set, and the kernel itself is a reduction, at this time the total space complexity is $O(1)$; the worst situation is that algorithm needs to judge $m - i$ times, at this time, the process of calculating $\underline{\tilde{B}}_{(\alpha,\beta,k)}^{\wedge}(\tilde{X})$ needs to be looped $m - i$ times, so at this time the total space complexity is $O(m)$.

In summary, at the best of times, the total time complexity is $O(m^2n^2)$, and the total space complexity is $O(1)$; at the worst of times, the total time complexity is $O(m^3n^2)$, and the total space complexity is $O(m)$.

The reduction algorithm ideas of the optimistic double-quantitative models are roughly the same as the model based on the $\tilde{B}_{(\alpha,\beta,k)}^\wedge(\tilde{X})$ invariant reduction algorithm. The difference is only that the function of the lower approximation operator is different. We Only need to be replace the lower approximation function in the above algorithm with optimistic double-quantitative model's lower approximation function, so we will not go into details at here.

In the following, we will combine an example to reduce the attributes of the four rough set models proposed earlier in this paper, and explore the characteristics of each model.

Example 5.1. *Given a target set-valued information system $I = (U, A, V, D, F)$ (Table 2). The universe of discourse $U = \{x_1, x_2, \dots, x_9\}$, the conditional attribute set $A = \{a_1, a_2, \dots, a_4\}$, Let $\alpha=0.68, \beta=0.45, k=0.72$. We want to find the reduction results of the four rough set models in the set-valued information system I .*

Table 2 A decision set-valued information system

U	a_1	a_2	a_3	a_4	d
x_1	{19, 4.4}	{3.2, 0.2}	{1.3, 5}	{1}	{1}
x_2	{28, 5.8}	{2.7, 1.2}	{3.9, 3}	{1}	{2}
x_3	{4.5}	{2.3, 0.3}	{1.3, 5}	{0.8}	{1}
x_4	{61, 7.1}	{3.0, 2.1}	{5.9, 0}	{1}	{3}
x_5	{5.5}	{2.3, 1.3}	{4.0, 2, 0.1}	{1}	{2}
x_6	{60, 59}	{3.0, 1.8}	{5.1, 1}	{1.1}	{3}
x_7	{51, 7.0}	{3.2, 1.4}	{4.7, 0}	{1}	{2}
x_8	{31, 4.8}	{3.1, 0.2}	{1.6, 5}	{1}	{1}
x_9	{5.7}	{3.0, 1.2}	{4.2, 4}	{0.8}	{2}

Since the method of computing the lower approximation is the same to Example 4.1, we will not go into details at here. According to the algorithm proposed, we separately calculate the reductions for four models and obtain the results in Table 3.

Table 3 Calculation results of the degree of change

Model	Lower approximation	Kernel element	Reduction
NB-SVP-SqRS	$\{a_2, a_3, a_4\}$	$\{a_1, a_2, a_3, a_4\}$	$\{a_1, a_2, a_3, a_4\}$
NB-SVD-SqRS	$\{x_2, x_3, x_4, x_7, x_8\}$	$\{a_1, a_2\}$	$\{a_1, a_2, a_4\}$
NB-SVO-DqRS	$\{x_2, x_3, x_4, x_7, x_8\}$	\emptyset	$\{a_2, a_3, a_4\}$
NB-SVP-DqRS	$\{x_2, x_4, x_7, x_8\}$	$\{a_3, a_4\}$	$\{a_2, a_3, a_4\}$

According to the results, the lower approximation under the set-valued information system of these four models is $\{x_2, x_3, x_4, x_7, x_8\}$ or $\{x_2, x_4, x_7, x_8\}$. Though, those models which have

same lower approximation results, their lower approximation reduction results are different, it is due to the different constraints of four models. Comparing the NB-SVP-SqRS model with the NB-SVD-SqRS model, the lower approximation and lower approximation reduction of these two models are both different. The reason is that although the two models have introduced quantitative information when they are constructed, the types of quantitative information are different, the constraints the models are classified are also different. When the model is classified, the former are focuses on relative quantitative information, while the latter are focuses more on absolute quantitative information. Then Comparing the NB-SVP-SqRS model with the NB-SVD-SqRS model, it is clear that the lower approximations of the two models are not the same but the attribute reduction results are the same, this is caused by the introduction of different logical operators in the two models. Compared with the optimistic double-quantitative model, the logical operators introduced by the pessimistic double-quantitative model make the lower approximation conditions of the model more stringent, so it has fewer elements, the lower approximation of optimistic double-quantitative rough set model is a subset of the pessimistic double-quantitative rough set model. Finally, compared with the NB-SVO-DqRS model and the NB-SVD-SqRS model, when the two models have the same lower approximation, but the reduction results are different. The reason is that, compared with the single-quantitative model, the constraint conditions of the approximation operators are increased in the double-quantitative model due to the introduction of logical operators, the requirements of the reduction are changed, and the reduction results will also be affected.

§6. Analysis on the advantages and disadvantages of quantitative rough set models

In order to clearly reflect the retention of the features contained in the original information system after the reduction of the four models, in this section, we use three data sets of different sizes as the background to calculate the lower approximation reduction results of the quantitative rough set models. Combining with the KNN algorithm, we can obtain the classification accuracy after reduction to reflect the classification ability of each model after reduction, and then we can compare the advantages and disadvantages of each model. Table 4 gives some basic information of the four experimental data tables.

Table 4 Basic information of the data sets

Name	Attribute	Object	Classification
segmentation	9	210	7
audit	13	776	2
HCV	14	1385	4

K nearest neighbor algorithm (K-Nearest neighbor, KNN) is one of the relatively simple and theoretically mature classification algorithms in machine learning. The basic idea is to

first calculate the K neighbors in the training sample set which are closest to the sample to be classified. If most of the K neighbor samples belong to a certain category, then the sample also belongs to this category. Then we will use the idea of KNN algorithm and replace the distance function of the traditional KNN classifier with the distance function constructed in Section 2. Based on the traditional KNN classifier, we redefine the applicable KNN classifier under the set-valued information system, short as KNN-SVIS. The data set is divided into training set and test set according to the ratio of 7:3. We calculate the lower approximation reduction of four models under each data set, then use the KNN-SVIS classifier to calculate the accuracy. The calculation results are shown in the following Table 5.

Table 5 Calculation results of accuracy rate

Data set	Model	Degree of change
segmentation	Attribute set before reduction	30.2%
	NB-SVP-SqRS	25.4%
	NB-SVD-SqRS	17.5%
	NB-SVO-DqRS	14.3%
	NB-SVP-DqRS	25.4%
audit	Attribute set before reduction	94.4%
	NB-SVP-SqRS	94.4%
	NB-SVD-SqRS	94.8%
	NB-SVO-DqRS	94.8%
	NB-SVP-DqRS	94.4%
HCV	Attribute set before reduction	29.1%
	NB-SVP-SqRS	29.1%
	NB-SVD-SqRS	26.0%
	NB-SVO-DqRS	26.0%
	NB-SVP-DqRS	29.1%

It can be seen from the calculation results shown in Table 5 that the performance of the four models is different under different data sets. In the first information system, the NB-SVP-SqRS model and NB-SVP-DqRS performed better, while the classification accuracy rate of the NB-SVD-SqRS model reduction is significantly lower. In the second information system, the NB-SVD-SqRS model and the NB-SVO-DqRS model performed better. In the third information system, the NB-SVP-SqRS model and NB-SVP-DqRS model perform better. Due to the different restriction conditions of the four models, the applicable models are different. NB-SVP-SqRS model and NB-SVP-DqRS model are more suitable for data sets "segmentation" and "HCV", while NB-SVD-SqRS model and the NB-SVO-DqRS model are more suitable for the data set "audit". By comparing the four models horizontally, the reduction results of the NB-SVP-SqRS model and the NB-SVP-DqRS model have higher classification accuracy than the NB-SVD-SqRS model and the NB-SVO-DqRS model under these three data sets, they have

stronger ability to retain the characteristics of the original information system.

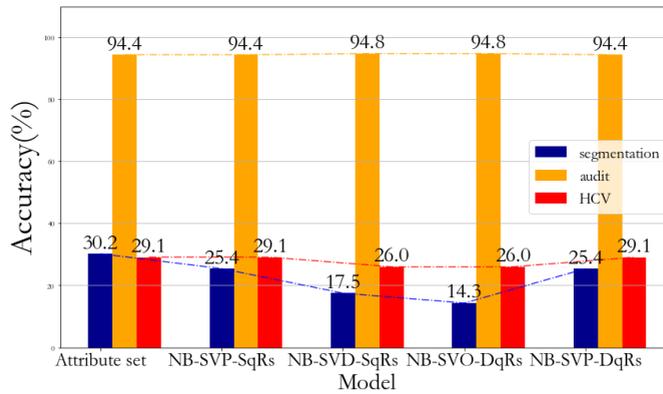


Fig. 1 The classification accuracy for different data sets.

It is clear from Figure 1 that except for the data set "audit", the classification accuracy rates of the other two data sets are unsatisfactory, the classification accuracy rates of these four models on the data sets "segmentation" and "HCV" are all lower than 30%. If only judge the model from the calculation results of the accuracy rate, these four models are not suitable for knowledge mining on the data sets "segmentation" and "HCV", but such conclusions are obviously one-sided, because the KNN-SVIS model has model errors, even the original attribute set is not 100% accuracy when using the KNN-SVIS model for classification prediction, the classification accuracy of the reduction result will also be affected by the model error. If only the size of the accuracy rate calculation result is considered, it cannot truly reflect the feature retention of the original information system by the attribute reduction result of the model.

In order to eliminate the influence of the KNN model error and truly reflect the change of the information contained in the reduction result compared with the original information system, the concept of the degree of change is proposed here.

Suppose P_A is the accuracy rate of the original attribute set after KNN-SVIS model reconstruction, and P_a is the accuracy rate of a model reduction result after KNN-SVIS model reconstruction, then define the degree of change, $D = \frac{P_a - P_A}{P_A}$.

Obviously, $D \in (-1, 1)$, the smaller D indicates that the reduction result retains the fewer features of the original information system.

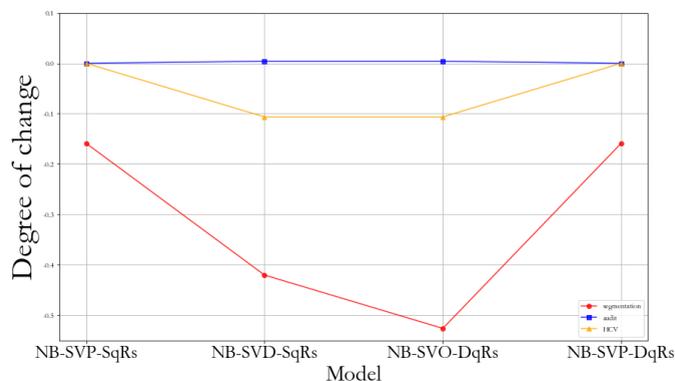
According to the above formula, the degree of change of the four models under the three data sets are calculated respectively. The calculation results are shown in Table 6.

It can be seen from Table 6 that in the first information system, the attribute reduction results of the NB-SVP-DqRS model and the NB-SVP-SqRS model are better to retain the characteristics of the original information system, compared with the original information system, in those model, only about 15% of the features are lost, while the reduction results of the NB-SVD-SqRS model and the NB-SVO-DqRS model respectively lose about 42% and 53%

Table 6 Calculation results of the degree of change

Data set	Model	Degree of change
segmentation	NB-SVP-SqRS	-0.1589
	NB-SVD-SqRS	-0.4205
	NB-SVO-DqRS	-0.5265
	NB-SVP-DqRS	-0.1589
audit	NB-SVP-SqRS	0
	NB-SVD-SqRS	0.042
	NB-SVO-DqRS	0.042
	NB-SVP-DqRS	0
HCV	NB-SVP-SqRS	0
	NB-SVD-SqRS	-0.1065
	NB-SVO-DqRS	-0.1065
	NB-SVP-DqRS	0

of the features. In the second information system, the degree of change D of the NB-SVD-SqRS model and the NB-SVO-DqRS model are both greater than 0, it indicates that the attribute reduction can more accurately describe the characteristics of the information system than the original attribute set, compared with the original information system, the characteristics of the information system contained in the reduction results of the NB-SVP-DqRS model and the NB-SVP-SqRS model remain unchanged. In the third information system, the attribute reduction results of the NB-SVD-SqRS model and the NB-SVO-DqRS model retain the characteristics of the original information system poorly, compared with the original information system that only lost about 11% of the features, the reduction results of the NB-SVP-DqRS model and the NB-SVP-SqRS model did not change the features of the information system contained in the original information system.

**Fig. 2** The degree of change under different data sets.

It shows that these three information systems can all use the above rough set model for knowledge mining, but due to the differences between different information systems, different rough set models are applicable to different information systems. The data set segmentation and HCV are more suitable to use NB-SVP-DqRS model and NB-SVP-SqRS model for knowledge mining, and as to the data set audit, NB-SVD-SqRS model and NB-SVO-DqRS model are more suitable.

In a word, for different information systems, the NB-SV-SqRS model and the NB-SV-DqRS model show different advantages. Due to the different restriction conditions of the lower approximation, the reductions of these four models also show different characteristics. When processing data, we can choose the model according to actual needs.

§7. Concluding remarks

This paper proposes single-quantitative rough set models and double-quantitative rough set models in the set-valued information system, and the related lower approximation reduction algorithms for neighborhood-based set-valued double-quantitative rough set model are also provided. Moreover, respectively calculate the lower approximation reduction of probability single-quantitative, degree single-quantitative, optimistic double-quantitative and pessimistic double-quantitative rough set models are discussed based on the proposed quantitative rough set models, and then the lower approximation reduction algorithms are carefully investigated. Finally, the KNN algorithm was applied to evaluate the accuracy of the classification of the proposed models after lower approximation reduction. In the future work, we will focus on our model as a classifier, and compare it with other existing classical classifiers to obtain its classification effect.

[References]

- [1] CHEN D, YANG Y. Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set models[J]. *IEEE Trans. Fuzzy Syst.*, 2014, 22(5): 1325-1334.
- [2] COOK W D, KRESS M, SEIFORD LM. A general framework for distance-based consensus in ordinal ranking models[J]. *Eur. J. Oper. Res.*, 1997, 96(2): 392-397.
- [3] FAN B, TSANG E C C, XU W, YU J. Double-quantitative rough fuzzy set based decisions: A logical operators method[J]. *Inf. Sci.*, 2017, 378: 264-281.
- [4] KHALIFEH M H, YOUSEFI-AZARI H, ASHRAFI A R, et al. Some new results on distance-based graph invariants[J]. *Eur. J. Combin.*, 2009, 30(5): 1149-1163.
- [5] LI W, XU W. Double-quantitative decision-theoretic rough set[J]. *Inf. Sci.*, 2015, 316: 54-67.
- [6] LINGRAS P, CHEN M, MIAO D. Qualitative and quantitative combinations of crisp and rough clustering schemes using dominance relations[J]. *Int. J. Appr. Reason.*, 2014, 55: 238-258.
- [7] LIU Y, LU Y. Attribute reduction in set-valued information system based on similarity[J]. *J. Nanjing Uni. (Natur. Sci.)*, 2015, 51(02): 384-389.
- [8] MA J, ZHANG W X. Information quantity-based attribute reduction in set-valued information systems[J]. *Fuzzy Syst. Math.*, 2013, 27(02): 177-182.
- [9] PAWLAK Z. Rough set[J]. *Int. J. Comp. Inf. Sci.*, 1982, 11(5): 341-356.

- [10] PAWLAK Z , WONG S K M, ZIARKO W. Rough sets: probabilistic versus deterministic approach[J]. Int. J. Man Mach. Stud., 1988, 29(1): 81-95.
- [11] SUN B, GONG Z. Variable precision probabilistic rough set model[J]. J. NW Norm. Uni. (Natur. Sci. Ed.). 2005, 4: 23-26.
- [12] XU W, GUO Y. Generalized multigranulation double-quantitative decision-theoretic rough set[J]. Knowl. Based Syst., 2016, 105: 190-205.
- [13] YAO Y, ZHAO Y. Attribute reduction in decision-theoretic rough set models[J]. Inf. Sci., 2008, 178(17): 3356-3373.
- [14] YAO Y. Three-way decisions with probabilistic rough sets[J]. Inf. Sci., 2009, 180(3): 341-353.
- [15] ZHANG N, MIAO D, YUE X. Approaches to knowledge reduction in interval-valued information systems[J]. Comput. Res. Devel., 2010, 47(08): 1362-1371.
- [16] ZHANG W X, MI J S, WU W Z. Comparative studies of knowledge reductions in inconsistent systems[J]. Fuzzy Syst. Math., 2003, 3: 54-60.
- [17] ZHANG W X, MI J S, WU W Z. Knowledge reductions in inconsistent information systems[J]. J. Comput., 2003, 3: 54-60.
- [18] ZHANG X, MIAO D. An expanded double-quantitative model regarding probabilities and grades and its hierarchical double-quantitative attribute reduction[J]. Inf. Sci., 2015, 299: 312-336.
- [19] ZHANG X, MIAO D. Two basic double-quantitative rough set models for precision and graded and their investigation using granular computing[J]. Int. J. Appr. Reason., 2013, 54: 1130-1148.
- [20] ZHANG X, MO Z, XIONG F, CHENG W. Comparative study of variable precision rough set model and graded rough set model[J]. Int. J. Appr. Reason., 2012, 53(1): 104-116.
- [21] ZIARKO W. Variable precision rough sets model[J]. J. Comput. Syst. Sci., 1993, 46(1), 39-59.