Postprint: Information Table Dimensionality Reduction Based on Knowledge Partition under Neighborhood Rough Sets

Authors: Peng Xiaoran, Liu Zunren, Ji Jun

Date: 2018-05-20T00:00:00+00:00

Abstract

Knowledge reduction in Pawlak rough sets encompasses knowledge reduction for both decision tables and information tables. As an extension of Pawlak rough sets, neighborhood rough sets are widely applied to attribute reduction for decision tables, yet rarely applied to attribute reduction for information tables. To design an attribute reduction algorithm suitable for information tables, we first propose an information table knowledge reduction criterion for neighborhood rough sets based on the information table knowledge reduction criterion of Pawlak rough sets. Subsequently, according to this criterion and combined with a greedy strategy, we further propose an information table attribute reduction algorithm for clustering tasks. Compared with the principal component analysis (PCA) algorithm, experimental results demonstrate that after dimensionality reduction of datasets using this algorithm, the obtained attribute reduction set contains a relatively larger number of attributes, and the K-means algorithm achieves higher clustering accuracy based on the attribute set. Experimental results prove that the algorithm can be effectively applied to attribute reduction for information tables.

Full Text

Preamble

Dimension Reduction for Information Tables Based on Knowledge Partition of Neighborhood Rough Set

Peng Xiaoran[1], Liu Zunren[2], Ji Jun[2]
([1]College of Data Science & Software Engineering, [2]College of Computer Science & Technology, Qingdao University, Qingdao, Shandong 266071, China)

Abstract: Knowledge reduction in Pawlak rough set theory encompasses two aspects: knowledge reduction for decision tables and knowledge reduction for information tables. As an extension of Pawlak rough sets, neighborhood rough sets have been widely applied to attribute reduction for decision tables, yet rarely for information tables. To design an attribute reduction algorithm suitable for information tables, this paper first proposes a knowledge reduction criterion for information tables in neighborhood rough sets, based on the knowledge reduction criterion for information tables in Pawlak rough sets. Subsequently, according to this criterion and incorporating a greedy strategy, we propose an attribute reduction algorithm for information tables applicable to clustering tasks. Compared with the Principal Component Analysis (PCA) algorithm, experimental results demonstrate that after dimension reduction using the proposed algorithm, the resulting attribute reduction sets contain more attributes, and the clustering accuracy of K-means algorithm based on these attribute sets is higher. The experimental results prove that the proposed algorithm can be effectively applied to attribute reduction for information tables.

Key Words: dimension reduction; clustering; information table; neighborhood rough set; attribute reduction

0 Introduction

Clustering is the process of grouping samples with similar properties and characteristics into the same category. With the rapid development of the information age, clustering faces not only the "data explosion" problem but also the more critical "curse of dimensionality" phenomenon caused by high-dimensional data [1]. Therefore, deleting redundant knowledge from datasets and eliminating interference from noisy data to the greatest extent possible, while maintaining knowledge representation capability unchanged, is highly meaningful for improving clustering algorithm efficiency.

Knowledge reduction constitutes an important research area in rough set theory. Pawlak rough set knowledge reduction includes reduction for both decision tables and information tables, with the primary distinction being the presence or absence of decision attributes. Classical Pawlak rough sets [2] process data through equivalence partitioning. However, this equivalence partitioning is only applicable to discrete data, whereas real-world applications often involve numerical data. This limitation has hindered the application of rough set theory.

To address this limitation, Zadeh [3] proposed the concepts of information granulation and granular computing. Lin [4] introduced the neighborhood model based on information granulation and granularity. Hu et al. [5] further proposed the neighborhood rough set model [6,7] based on neighborhood granulation and rough approximation concepts, which can handle numerical data and significantly expands the application scope of rough set theory.

As an extension of Pawlak rough sets, neighborhood rough sets have been widely applied to attribute reduction for decision tables [5,8-12]. For instance, Hu [5] proposed a forward greedy-based attribute reduction algorithm for decision tables; Liu [8] combined particle swarm optimization to propose a fast reduction algorithm for high-dimensional datasets; Liu [9] proposed a fast decision table attribute reduction algorithm based on mapping partitioning, among others. These algorithms rely on positive region computation in neighborhood rough sets, which is based on decision attributes—i.e., when sample categories are known. Consequently, attribute reduction algorithms designed for decision tables are not applicable to information tables.

Currently, neighborhood rough sets have rarely been applied to attribute reduction for information tables. To extend the application of neighborhood rough sets to information tables and design an attribute reduction algorithm for information tables suitable for clustering tasks, this paper first proposes a knowledge reduction criterion for numerical information tables based on Pawlak rough set criteria for information table reduction. Building upon this criterion and incorporating greedy strategy, we further propose a forward greedy-based information table attribute reduction algorithm (Fast Attribute Reduction Algorithm for Information Table, FARAIT). Finally, through comparison with PCA—an unsupervised learning algorithm—experimental verification demonstrates the effectiveness of the proposed algorithm.

---

1 Related Concepts

1.1 Knowledge Partition Under Pawlak Rough Sets [2]

Classical Pawlak rough sets consider knowledge to be granular, representing a capability to classify samples in a universe.

Definition 1 (Information Granule). Let U be a non-empty finite set of samples, called the universe. Any subset $X \subseteq U$ of the universe is called a concept or category of U, where each concept represents an information granule of U.

Definition 2 (Indiscernibility Relation). Given a universe U and a cluster of equivalence relations P on U, if $P \subseteq P$ and $P \neq \emptyset$, then $IND(P) = \bigcap_{R \in P} R$ (the intersection of all equivalence relations in P) remains an equivalence relation on U, called the indiscernibility relation on U, denoted as $IND(P)$.

Based on the above definitions, for a knowledge base $K = (U, P)$, where P is a cluster of equivalence relations on U, if $P \subseteq P$ and $P \neq \emptyset$, then $U/IND(P) = \{[x]_{IND(P)} \mid x \in U\}$ represents the knowledge related to P, i.e., the universe U is partitioned into several equivalence classes (information granules) according to the equivalence relation $IND(P)$. For example, if $U = \{x_1, x_2, x_3, x_4, x_5\}$, then U can be partitioned according to relation E_1 as $\{\{x_1\}, \{x_2\}, \{x_3, x_4, x_5\}\}$.

1.2 Knowledge Partition Under Neighborhood Rough Sets [5]

Pawlak rough sets facilitate granular computing through equivalence relations, which can be directly constructed in discrete datasets but not in numerical datasets. As an extension of Pawlak rough sets, neighborhood rough sets have been effectively applied to processing numerical datasets.

Definition 3 (Metric Computation). Given an n-dimensional real number space $\mathbb{R}^n$, for any two points $x_i = (x_{i1}, x_{i2}, ..., x_{in})$ and $x_j = (x_{j1}, x_{j2}, ..., x_{jnn})$ in the space, define the metric computation as:

$$d(x_i, x_j) = \sqrt{\sum_{p=1}^{n} (x_{ip} - x_{jp})^2}$$

which constitutes a metric on $\mathbb{R}^n$.

Definition 4 (Neighborhood Granulation). In real number space, let $U = \{x_1, x_2, ..., x_n\}$ be a non-empty finite set of samples. For any $x_i \in U$ and $\delta \geq 0$, define:

$$\delta(x_i) = \{x_j \mid x_j \in U \wedge d(x_i, x_j) \leq \delta\}$$

as the δ-neighborhood information granule generated by x_i, simply called the neighborhood granule of x_i.

Based on the above definition, for a knowledge base $K = (U, B)$, where $B \subseteq B$ and $B \neq \emptyset$, the universe U can be partitioned into several neighborhood information granules according to relation B. For example, if $U = \{x_1, x_2, x_3, x_4, x_5\}$, then U can be partitioned according to relation E_2 as $\{\{x_1, x_2\}, \{x_2, x_3, x_4\}, \{x_3, x_5\}, \{x_4, x_5\}, \{x_5, x_1\}\}$.

1.3 Knowledge Representation System of Rough Sets [2]

Definition 5 (Knowledge Representation System). A quadruple $KRS = (U, A, V, f)$ is called a knowledge representation system, where: - $U = \{x_1, x_2, ..., x_n\}$ is a non-empty finite set of samples; - A is a non-empty finite set of attributes; - $V = \bigcup_{a \in A} V_a$ is the value domain of all attributes, where V_a is the value domain of attribute a; - $f : U \times A \to V$ is an information function.

In a knowledge representation system KRS, let $C, D \subseteq A$ be condition attributes and decision attributes, respectively. If $D \neq \emptyset$, the knowledge representation system is collectively called a decision table (DT); otherwise, it is called an information table (IT). Generally, decision tables are used for classification tasks, while information tables are used for clustering tasks.

Definition 6 [2] (Pawlak Rough Set Information Table Knowledge Reduction Criterion). Given an information table $S = (U, C \cup D, V, f)$, if the partition of universe U according to attribute set C is inconsistent with the partition according to attribute set $C - \{a\}$, then attribute a is called a non-deletable attribute in C; otherwise, a is called a deletable attribute in C.

2 Proposed Information Table Knowledge Reduction Criterion

Deleting redundant knowledge from a knowledge system while maintaining its knowledge representation capability unchanged is called knowledge reduction. For an information table, different knowledge corresponds to different collections of condition attributes, and redundant knowledge refers to deletable attributes.

According to Definitions 2 and 4, knowledge partitioning under neighborhood rough sets differs from that under Pawlak rough sets. Compared with Pawlak rough sets, neighborhood rough sets transform the equivalence relation on the universe into a covering relation. If Definition 6 is directly applied to knowledge partitioning under neighborhood rough sets, since each sample in the universe forms a neighborhood information granule and each granule contains different samples, the comparison process for determining whether two partitions in Definition 6 are consistent becomes inconvenient and computationally expensive. To address this issue and considering the characteristics of knowledge partitioning under neighborhood rough sets, this paper proposes an equivalent transformation of knowledge partition changes through Theorem 1.

Theorem 1. For an information table, given a certain δ value, during the knowledge reduction process in neighborhood rough sets, determining whether the knowledge partition of the universe is consistent can be transformed into determining whether the number of mutually neighboring sample pairs is consistent.

Proof. For a numerical information table as shown in , according to the full attribute set C, the universe U can be mapped into a set of n m-dimensional vectors in real number space, i.e., $\{value_{i1}, value_{i2}, ..., value_{im}\}_{i=1}^{n}$. Similarly, according to a partial attribute set $B \subseteq C$, it can also be mapped into n t-dimensional vectors, where $1 \leq t \leq m$.

According to the concept of neighborhood granulation, the universe U can obtain a corresponding knowledge partition based on attribute set C, i.e., a collection of neighborhood information granules. Based on attribute set B (where each vector simultaneously reduces the dimension corresponding to $C-B$), a new knowledge partition can also be obtained. The key to determining whether attribute a is deletable lies in judging whether these two knowledge partitions are consistent. It can be seen that knowledge reduction is essentially a dimensionality reduction process in real number space.

For illustration, let n m-dimensional vectors be represented by 5 vectors, i.e., $U = \{x_1, x_2, x_3, x_4, x_5\}$, use two-dimensional space to represent m-dimensional high-dimensional space, and use one-dimensional space to represent the reduced t-dimensional low-dimensional space, thereby analyzing the knowledge partition change process during reduction.

As shown in [Figure 1: see original paper], dashed circles represent neighborhoods, where (a) and (b) denote high-dimensional space, and (c) denotes low-dimensional space. In the high-dimensional space shown in (a), U is partitioned as $\{\{x_1, x_2\}, \{x_2, x_3, x_4\}, \{x_3, x_5\}, \{x_4, x_5\}, \{x_5, x_1\}\}$, with 3 mutually neighboring sample pairs: $\{(x_1, x_2), (x_3, x_4), (x_4, x_5)\}$. Then, without considering the dimension corresponding to the attribute to be evaluated for deletion, the low-dimensional space and corresponding partition of U shown in (c) are obtained. Comparing (a) and (c), the partitions are consistent, and the number of mutually neighboring sample pairs remains the same, indicating that this attribute is deletable. If the high-dimensional space is as shown in (b), U is partitioned as $\{\{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_4, x_5\}, \{x_5, x_1\}\}$, with 2 mutually neighboring sample pairs: $\{(x_1, x_2), (x_4, x_5)\}$. Comparing (b) and (c), the partitions are inconsistent, and the number of mutually neighboring sample pairs has increased compared to the former. This indicates that the attribute is non-deletable.

Taking sample x_2 as an example to analyze the dimensionality reduction process: During reduction, sample points originally belonging to the circle (neighborhood) of x_2 in high-dimensional space will certainly remain within the circle in the reduced low-dimensional space because their Euclidean distance to x_2 further decreases after reducing one dimension. Sample points originally outside the circle may either remain outside or move inside the circle after reduction due to decreased Euclidean distance. For the number of mutually neighboring sample pairs, only two scenarios exist: remaining unchanged or increasing, corresponding respectively to attributes being deletable or non-deletable. This demonstrates that for a given δ value, determining knowledge partition consistency in neighborhood rough set reduction can be transformed into determining consistency in the number of mutually neighboring sample pairs.

Relatively speaking, calculating the number of mutually neighboring sample pairs is easier and can be accomplished using upper or lower triangular matrices. Based on Theorem 1, this paper obtains a more intuitive and convenient basis for judging knowledge partition changes, leading to the proposal of a knowledge reduction criterion for information tables in neighborhood rough sets.

Standard 1 (Neighborhood Rough Set Information Table Reduction Criterion). Given an information table $S = (U, C \cup D, V, f)$, based on neighborhood rough sets, let n be the number of mutually neighboring sample pairs in the neighborhood information granules partitioned by attribute set C in real number space, and n' be the number of mutually neighboring sample pairs partitioned by attribute set $C - \{a\}$. If $n = n'$, i.e., the partitions are consistent, then attribute a is deemed deletable in C. If $n \neq n'$, i.e., the partitions are inconsistent, then attribute a is deemed non-deletable in C.

3 Forward Greedy Information Table Attribute Reduction Algorithm

Definition 7 [1] (Attribute Reduction). Given an information table $S = (U, C \cup D, V, f)$, if the partition of universe U according to attribute set $B \subseteq C$ is consistent with the partition according to attribute set C, and any attribute $a \in B$ is non-deletable, then B is called an attribute reduction of C.

According to Definition 7, the objective of attribute reduction algorithms for information tables is to find a subset of attributes that yields the same knowledge partition as the original attribute set.

Based on Definition 7 and Standard 1, a "blind deletion method" is a straightforward approach. Its idea is: for an original attribute set of a dataset, select one attribute for deletion evaluation; if deletable, remove it from the attribute set and repeat the selection and evaluation on the new attribute set; if not deletable, evaluate the remaining attributes. The termination condition is when all attributes in the set are non-deletable.

Analysis of Blind Deletion Method Complexity. Under this algorithm, assuming a dataset has n samples and m attributes, with the reduction result containing k attributes, the best case occurs when the first $m-k$ attributes are deletable, while the worst case occurs when the last $m-k$ attributes are deletable. The time complexity can be expressed as:

- Best case: $1 \cdot n + 2 \cdot n + \cdots + (m-k) \cdot n + (m-k+1) \cdot n$
- Worst case: $(k+1) \cdot n + (k+2) \cdot n + \cdots + m \cdot n$

For an attribute set with m attributes, evaluating whether a particular attribute can be deleted requires computing knowledge partitions for n samples in m-dimensional real number space. The best case requires only one judgment per deletion, while the worst case requires m judgments per deletion.

The blind deletion method is a dimensionality reduction process from m dimensions to k dimensions in real number space. Since the initial dimensionality is high and k is typically much smaller than m for a given dataset, the computational cost is substantial when m is large and k is small. If this process is improved to an upward dimensionality process from 0 dimensions to k dimensions, the computational load would be significantly reduced.

Based on this insight and drawing inspiration from the forward greedy algorithm in [5], this paper proposes a forward greedy information table attribute reduction algorithm (Fast Attribute Reduction Algorithm for Information Table, FARAIT) to obtain an optimal or near-optimal attribute reduction for information tables according to Standard 1 in Section 2.

Definition 8 (Attribute Importance). Given an information table $S = (U, C \cup D, V, f)$ and $B \subseteq C$, the importance of attribute $a \in C - B$ relative to

B is defined as:

$$SIG(a, B) = Card(B) - Card(B \cup \{a\})$$

where $Card(B)$ denotes the number of mutually neighboring sample pairs in the neighborhood information granules formed by attribute set B.

When adding attributes to the reduction set, the upward dimensionality process in real number space increases distances between samples, reducing the number of mutually neighboring sample pairs. This reduction amount measures the importance of each attribute. Attribute importance can be understood as the capability to distinguish samples through that attribute. For example, an information table can be clustered into "male" and "female" categories based on the original attribute set. During greedy selection, the attributes "height" and "face shape" have different importance levels. In real number space, the distribution of "face shape" values is relatively uniform, while "height" values gradually cluster toward the middle. Comparatively, adding the "height" attribute to the reduction set reduces more mutually neighboring sample pairs, indicating stronger distinguishing capability and higher attribute importance.

Unlike the blind deletion method, forward search algorithms ensure that important attributes are added first, thus avoiding loss of critical attributes—a result difficult to guarantee with blind deletion. For information tables with numerous redundant attributes, deleting important attributes may not necessarily reduce the overall system' s distinguishing capability, potentially resulting in retention of many weakly distinguishing attributes that collectively maintain the original data' s 分辨能力, rather than a few highly distinguishing attributes.

The specific strategy of FARAIT is as follows: Initialize the attribute reduction set as empty, compute the importance of candidate attributes not in the reduction set each iteration, select the attribute with maximum importance to add to the reduction set, and continue until the importance of all remaining attributes becomes 0. At this point, the number of mutually neighboring sample pairs in the neighborhood information granules formed by the reduction set no longer changes, indicating that the corresponding knowledge partition is consistent with or similar to that formed under the original attribute set, and any remaining attributes are deletable according to Standard 1. The algorithm is shown as Algorithm 1.

Algorithm 1: FARAIT

Input: Information table $S = (U, C \cup D, V, f)$

Output: Attribute reduction red

1. Initialize $red = \emptyset$
2. If $Card(red) = 0$ // Here $Card(\emptyset) = n(n-1)/2$
3. For any $a_i \in C - red$

4. `Compute importance: $SIG(a_i, red) = Card(red) - Card(red \cup \{a_i\})$`
5. Select a_k satisfying: $SIG(a_k, red) = \max_i SIG(a_i, red)$
6. If $SIG(a_k, red) > 0$
7. `$red \leftarrow red \cup \{a_k\}$`
8. `Go to Step 2`
9. Else
10. Go to Step 11
11. Return red

Assuming a dataset has n samples and m attributes, with the reduction result containing k attributes, the time complexity of FARAIT can be expressed as:

$$n \cdot (m + (m-1) + \cdots + (m-k+1)) = n \cdot \left(k \cdot m - \frac{k(k-1)}{2} \right)$$

Evidently, due to the greedy strategy, FARAIT can obtain an optimal or near-optimal attribute reduction in relatively short time.

---

4 Experimental Analysis

As an effective unsupervised dimensionality reduction algorithm, PCA has been widely applied in artificial intelligence, pattern recognition, image processing, and other fields [13-15]. The experiments first discuss the impact of δ values on FARAIT to determine appropriate δ values. Then, under these δ values, both FARAIT and PCA are used for dimensionality reduction on datasets, followed by K-means clustering on the reduced datasets. Finally, the number of attributes in the reduction sets and K-means clustering accuracy are compared. K-means is used to evaluate the dimensionality reduction effectiveness of both algorithms, with the initial number of cluster centers set to the number of classes provided in the dataset.

4.1 Experimental Environment

The UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/) provides a series of standard datasets for testing. This paper selects 7 numerical datasets from UCI, each providing conditional attributes and decision attributes. The experiments are conducted on a PC with Intel(R) Core(TM) i5 CPU and 4GB RAM, using MATLAB R2016b on Windows 7.

shows the dataset descriptions.

4.2 FARAIT Algorithm Experimental Analysis

This section provides a detailed analysis of FARAIT. Without considering decision attributes in the datasets, FARAIT is first applied to reduce the dimensions of datasets wine, WDBC, and sonar under different δ values, followed by K-means clustering on both reduced and original datasets. The number of attributes in the reduction sets is recorded, and clustering results are compared with the decision attributes provided in the datasets to count correctly clustered samples and calculate accuracy (precision), thereby determining appropriate δ values.

4.2.1 δ Value and Attribute Reduction Count According to Definition 4, the δ value directly affects attribute reduction results. Different δ values yield different attribute reductions, resulting in different clustering accuracies. This paper selects 51 δ values in the interval $[0, 1]$ with increments of 0.02, recording the corresponding attribute reduction counts and clustering accuracies for each δ value. K-means is executed 20 times, with final clustering accuracy averaged.

The relationship between δ values and attribute reduction counts is shown in [Figure 2: see original paper]. For the same dataset, different δ values produce different attribute reduction counts. Starting from $\delta = 0$, as δ increases, the number of attributes in the reduction set increases until it stabilizes at the length of the original attribute set, at which point the δ value is called the saturation point.

4.2.2 δ Value and K-means Clustering Accuracy The relationship between δ values and K-means clustering accuracy is shown in Figure 3: see original paper-(c). The horizontal line represents clustering accuracy using the original attribute set, the polyline represents clustering accuracy using attribute reductions under different δ values, and the dashed line represents the saturation point. Using the saturation point as a benchmark, Figure 3: see original paper-(c) are divided into two parts: the first part corresponds to attribute reductions under different δ values, and the second part corresponds to the original attribute set.

Analysis of the second part: In all figures, the polyline fluctuates near the horizontal line, indicating that initial point selection affects K-means stability. The fluctuation degree relative to the horizontal line represents K-means stability for each dataset. In (a), the polyline fluctuates slightly, indicating relatively stable clustering on wine; in (b), the polyline perfectly matches the horizontal line, indicating extremely stable clustering on WDBC; in (c), the polyline fluctuates significantly because the attribute reduction length for sonar has not yet stabilized under the selected δ values.

Analysis of the first part: First, as δ increases, all polylines initially rise above then fall below the horizontal line, indicating that each dataset contains deletable redundant attributes that reduce clustering accuracy. In (a), the poly-

line starts below the horizontal line and increases with δ, indicating few redundant attributes in wine. In (b), at small δ values, the polyline is far above the horizontal line then drops rapidly, indicating many redundant attributes in WDBC that severely affect clustering accuracy. In (c), the polyline remains above the horizontal line with clear decreasing trend, indicating many redundant attributes in sonar that affect clustering accuracy.

4.2.3 FARAIT Experimental Conclusions The above analysis shows that with appropriate δ values, the forward greedy information table attribute reduction algorithm can effectively delete redundant attributes, eliminate their interference on clustering accuracy, and optimize clustering performance. Considering K-means instability, K-means clustering is relatively ideal when δ is in $[0.14, 0.18]$. Under these δ values, the obtained attribute reduction counts are smaller than the original datasets while achieving higher clustering accuracy.

4.3 Comparison Between FARAIT and PCA

Based on the experimental conclusions in Section 4.2, FARAIT is configured with $\delta = 0.16$ and PCA with a threshold of 0.85. Both dimensionality reduction algorithms are applied to the datasets, followed by K-means clustering. The number of attributes in reduction sets and clustering accuracy are recorded, with K-means executed 20 times and final accuracy averaged. Results are shown in .

shows the experimental comparison between FARAIT and PCA.

Based on , line charts of attribute reduction counts and K-means clustering accuracy are plotted in [Figure 4: see original paper] and [Figure 5: see original paper].

In [Figure 4: see original paper], the line representing the original attribute set is at the top, FARAIT in the middle, and PCA at the bottom. This indicates both FARAIT and PCA effectively reduce attribute counts. However, FARAIT produces more attributes in the reduction set compared to PCA.

In [Figure 5: see original paper], the FARAIT line is above the PCA line, indicating that K-means achieves higher clustering accuracy after dimensionality reduction with FARAIT compared to PCA.

In summary, compared with PCA, FARAIT yields more attributes in reduction sets and higher K-means clustering accuracy.

---

5 Conclusion

To extend the application of neighborhood rough sets to information tables, this paper designs an unsupervised information table dimensionality reduction algorithm suitable for clustering tasks. When processing numerical information

tables, this algorithm can be used for data preprocessing to delete redundant information, maintain or even improve clustering accuracy, and optimize clustering algorithm performance. The experiments demonstrate that δ values directly affect FARAIT' s effectiveness—appropriate δ values optimize FARAIT' s performance, while inappropriate values yield mediocre or poor results. Future work will investigate the optimal δ values for FARAIT with different clustering algorithms to achieve superior clustering performance.

References

[1] He L, Cai Y C, Yang Z. Survey of clustering methods for high-dimensional data [J]. Application Research of Computers, 2010, 27(1): 23-26.

[2] Pawlak Z, Słowinski R. Rough set approach to multi-attribute decision analysis [J]. European Journal of Operational Research, 1994, 72(3): 443-459.

[3] Zadeh L A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic [J]. Fuzzy Sets & Systems, 1997, 90(90): 111-127.

[4] Lin T Y. Granular Computing on binary relations I: Data mining and neighborhood systems [J]. Rough Sets in Knowledge Discovery, 1998(2): 107-121.

[5] Hu Q, Yu D, Liu J, Wu C. Neighborhood rough set based heterogeneous feature subset selection [J]. Information Sciences, 2008, 178(18): 3577-3594.

[6] Wang G Y. Rough Set Theory and Knowledge Acquisition [M]. Xi' an: Xi' an Jiaotong University Press, 2001: 147-156.

[7] Hu Q H, Yu D R. Applied Rough Computing [M]. Beijing: Science Press, 2012.

[8] Liu Z R, Wu G F. Fast reduction algorithm for high-dimensional datasets based on neighborhood rough model [J]. Computer Science, 2012, 39(10): 268-271.

[9] Liu Y, Huang W, Jiang Y, Zeng Z. Quick attribute reduct algorithm for neighborhood rough set model [J]. Information Sciences, 2014, 271(7): 65-81.

[10] Chen H, Li T, Cai Y, et al. Parallel Attribute Reduction in Dominance-based Neighborhood Rough Set [J]. Information Sciences, 2016, 373: 351-368.

[11] Wang C, Shao M, He Q, et al. Feature subset selection based on fuzzy neighborhood rough sets [J]. Knowledge-Based Systems, 2016, 111: 173-179.

[12] Chen Y, Zhang Z, Zheng J, et al. Gene selection for tumor classification using neighborhood rough sets and entropy measures [J]. Journal of Biomedical Informatics, 2017, 67: 59-68.

[13] Wang J, Feng J, Han Z Y. Fault detection based on locality preserving PCA algorithm using manifold learning [J]. Control and Decision, 2013(5): 683-687.

[14] Liu L M, Fan X P, Liao Z F, et al. A PCA dimension reduction algorithm based on L2,1-norm [J]. Application Research of Computers, 2013, 30(1): 39-41.

[15] Hosoya H, Hyvärinen A. Learning Visual Spatial Pooling by Strong PCA Dimension Reduction [J]. Neural Computation, 2016, 28(2): 1-16.

Note: Figure translations are in progress. See original paper for figures.