Fuzzy Systems and Data Mining VIII
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220398

Fast Attribute Reduction for Big Datasets Based on Neighborhood Rough Set

Cheng Wang [a], Yabin Shao [a,b,1], Xiaoli Peng [a,c],Shuyin Xia [a]and Weiqi Chen [a]

[a] *School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*

[b] *School of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*

[c] *Institute of Intelligent Manufacturing Technology, Sichuan University of Arts and Science, Dazhou, Sichuan 635000, China*

Abstract. Neighborhood rough set (NRS) is usually only applicable to small datasets due to the large number of useless and repetitive neighborhood calculations, which severely limits the efficiency of NRS. Many studies improve the efficiency of NRS by narrowing the neighborhood search range down and achieve good performance on small datasets, but they do not perform well on big datasets. To further improve the efficiency on big datasets, we propose a fast attribute reduction method for big datasets based on NRS (FARforBD). In addition, a theorem is also represented to prove the correctness and effectiveness of the proposed method. In FARforBD, we further reduce the neighborhood search range to a neighborhood without any positive region samples. This method greatly reduces many useless neighborhood calculations. The comparison experiments on big datasets show the effectiveness and efficiency of FARforBD.

Keywords. Attribute reduction, Neighborhood rough set, Big datasets, Fast neighborhood calculations.

1. Introduction

Attribute reduction, also called feature selection, is an important application of Pawlak's rough set theory [1,2]. The core idea of it is to remove redundant or irrelevant attributes from the condition attribute set while keeping the same discrimination ability as the originate attribute set. Thus, attribute reduction can both reduce the computing complexity and avoid the curse of dimensionality [3,4,5].

The classical attribute reduction algorithms are based on equivalence classes and can only be used for discrete data. But for continuous data, these algorithms must be used after discretization. However, discretization will lead to information losing[6,7], so the classical attribute reduction algorithms do not perform well on the continuous data. As a result, the generalized rough set appears, such as NRS[8,9,10,11,12], fuzzy rough set (FRS)[13,14,15,16], covering rough set (CRS)[17,18,19,20,21,22,23,24], and so on.

[1]Corresponding Author: Yabin Shao, Chongqing University of Posts and Telecommunications, E-mail shaoyb@cqupt.edu.cn.

NRS replaces the equivalence relation or partition by using neighborhood relation, which is measured by distance metric. Because of the simplicity and intuitiveness of processing continuous data, NRS is favored by many scholars. Hu et al. proposed some NRS based algorithms for continuous data [25] and mixed format data [26,27]. Sun et al. proposed the fuzzy dominant neighborhood rough set (FDNRS) for multi-label data [28]. Sang et al. proposed a method based on fuzzy dominance NRS for dynamic interval-valued data [29]. Li et al. proposed a multi-criterion approach for Neighborhood attribute reduction [8]. Although these NRS-based algorithms perform well in attribute reduction, they are low efficiency because of too many neighborhood calculations. The time complexity of neighborhood calculation in these NRS-based method is as high as $O(n^2)$. Hu et al. prospered a fast forward attribute reduction method (FARNeMF) to improve the efficiency of NRS [30]. Liu et al. proposed an efficient attribute reduction algorithm based on NRS by dividing the records of the whole dataset into many buckets [31]. Peng et al. proposed a fast neighborhood calculation framework (FNC) and applied it to NRS attribute reduction [32]. FNC is the fastest neighborhood calculation method we know so far.

FNC is indeed very efficient on small datasets, but it is still not good enough on large datasets. Thus, we propose FARforBD based on NRS.

2. Preliminaries

In this section, we review some theories and notions of NRS, which are used throughout our work in this study.

Definition 1 ([32]) *Given a nonempty finite set $U = \{x_1, x_2, ..., x_n\}$, where U is the universe; $A = \{a_1, a_2, ..., a_m\}$ is the attribute set of U; $V = \bigcup_{a \in A} V_a$ is the collection of attribute values, where V_a presents the value range of attribute a; $I = U \times A \to V$ is the mapping function between the sample and its corresponding attribute value. Then, $IS = \langle U, A, V, I \rangle$ is called an information system.*

Definition 2 ([32]) *Given an information system $IS = \langle U, A, V, I \rangle$, $A = C \bigcup D$, where C and $D(D \neq \emptyset)$ are the condition and decision attribute set. Then, $DS = \langle U, C, D \rangle$ is called a decision system.*

Definition 3 ([32]) *Given an m-dimensional real space R, there exists a mapping $\Delta : R^N \times R^N \to R$, where Δ is a metric on R. $\forall x, y, z \in R$, Δ satisfies:*

(1) *Positivity: $\Delta(x,y) \geq 0$, $\Delta(x,y) = 0$ if $x = y$;*
(2) *Symmetry: $\Delta(x,y) = \Delta(y,x)$;*
(3) *Triangle inequality: $\Delta(x,z) \leq \Delta(x,y) + \Delta(y,z)$.*

Then, $\langle R, \Delta \rangle$ is called a distance space or a metric space.

Δ is the distance, which is always expressed as L_p-norm:

$$\Delta_B(x, x_i) = (\sum_{j=1}^{s} \left| I(x, a_j) - I(x_i, a_j) \right|^p)^{\frac{1}{p}}, \tag{1}$$

where B is an attribute set and $a_j \in B$; s is the attribute number in B; $\Delta_B(x, x_i)$ is the distance between x and x_i with respect to B.

Definition 4 ([32]) *Given a decision system $DS = \langle U, C, D\rangle$, for all $x_i \in U$ and $B \subseteq C$, the $\delta-$neighborhood of x_i with respect to B is defined as follow*

$$\delta_B(x_i) = \{x | x \in U, \Delta_B(x_i, x) \leq \delta\}, \quad s.t. \quad \delta > 0. \tag{2}$$

As an important parameter in NRS, δ is the neighborhood radius and needs to be optimized.

Definition 5 ([32]) *Given a decision system $DS = \langle U, C, D\rangle$, $X_1, X_2, \ldots, X_k$ are the equivalence classes, which are the division of D to U. The lower approximation, upper approximation, positive region, negative region and boundary region of B related to D are as follows.*

$$\underline{N}D = \bigcup_{i=1}^{k} \underline{N}X_i, \tag{3}$$

$$\overline{N}D = \bigcup_{i=1}^{k} \overline{N}X_i, \tag{4}$$

$$POS(D) = \underline{N}D, \tag{5}$$

$$BN(D) = \overline{N}D - \underline{N}D, \tag{6}$$

where $\underline{N}X_i = \left\{x_j | x_j \in U, \delta_B(x_j) \subseteq X_i\right\}$; $\overline{N}X_i = \left\{x_j | x_j \in U, \delta_B(x_j) \cap X_i \neq \emptyset\right\}$.

Theorem 1 ([30,32]) *Given a decision system $DS = \langle U, C, D\rangle$, B_1 and B_2 are two attribute subsets of C. If $B_1 \subseteq B_2 \subseteq C$, then $POS(D)_{B_1} \subseteq POS(D)_{B_2}$.*

Theorem 2 ([30,32]) *Given a decision system $DS = \langle U, C, D\rangle$, $B_1 \subseteq B_2 \subseteq C$. If for all $x \in U$, there has $x \in POS(D)_{B_1}$, then $x \in POS(D)_{B_2}$.*

In classical NRS, we need to calculate the neighborhood relations among all samples in U. Thus, the time complexity of the classical NRS is $O(n^2)$. Hu et al. proposed Theorems 1 and 2 in FARNeMF to prove that when we add a new candidate attribute to B_1, there is no need to calculate the neighborhood relations among the samples which have been in the positive region [30]. That is to say, when adding new attributes to B_1, we only need to calculate the neighborhood relations among the boundary region samples. Therefore, the neighborhood search range is reduced from U to the boundary region. To further reduce the neighborhood search range, FNC propose the following definition and theorem.

Definition 6 ([32]) *Given a decision system $DS = \langle U, C, D\rangle$, B_1 and B_2 are two attribute subsets of C. If $B_1 \subset B_2$, we call B_1 the child attribute set and B_2 the parent attribute set.*

Theorem 3 ([32]) *Given a decision system $DS = \langle U, C, D\rangle$, $B_1 (B_1 \neq \emptyset)$ and B_2 are two attribute subsets of C. If $B_1 \subseteq B_2 \subseteq C$, then $\forall x \in U$, $\delta_{B_2}(x) \subseteq \delta_{B_1}(x)$.*

FNC uses Theorem 3 to prove that the neighborhood search range can be reduced to the neighborhood of the child attribute set. That is to say, when we have got $\delta_{B_1}(x)$, we can use it as the neighborhood search range of $\delta_{B_2}(x)$.

3. Fast Attribute Reduction for Big Datasets Based on Neighborhood Rough Set

In a big dataset, there may be lots of samples in $\delta(x_i)$, so the neighborhood search range still very large. In FNC, although the neighborhood search range is reduced from the boundary region to the neighborhood of the corresponding child attribute set, where there are some positive region samples. It still spends a lot of time to judge whether the positive region samples in the neighborhood of the corresponding child attribute set are in the neighborhood of the corresponding father attribute set. According to Theorem 3, $\delta_B(x)$ is the neighborhood search range of $\delta_{B\cup a_i}(x)(a_i \in (B-C), a_i \notin B)$. The positive region samples in $\delta_B(x)$ are calculated for judging whether they are still in $\delta_{B\cup a_i}(x)$. In fact, this kind of calculation and judgment is useless and unnecessary. In FARforBD, we delete the positive region sample in the neighborhood of the father attribute set. Therefore, we can delete the positive region sample in the neighborhood of the child attribute set. In order to prove the effectiveness of this theory, we give Theorem 4.

Theorem 4 *Given a decision system $DS = \langle U, C, D\rangle$, $B \subseteq C$. If $x_i \in POS(D)_B$ and $x_i \in \delta_B(x_j)$, then x_i can be deleted form $\delta_B(x_j)$.*

Proof: According to the symmetry of Δ in Definition 3, we get

$$\Delta(x_i, x_j) = \Delta(x_j, x_i). \tag{7}$$

Combining Eq. (7) with Definition 4 and Eq. (2), we have

$$x_i \in \delta_B(x_j), \tag{8}$$

$$x_j \in \delta_B(x_i). \tag{9}$$

Since $x_i \in POS(D)_B$, in terms of Definition 5, the labels in $\delta_B(x_i)$ are the same. That is to say,

$$label(x_i) = label(x_j), \tag{10}$$

where $label(x)$ denotes the label of x.

When adding any candidate attribute $a_i(a_i \in (C-B))$ to B, according to Theorem 4, $\delta_B(x_j)$ is the neighborhood search range of $\delta_{B\cup\{a_i\}}(x_j)$.

(1) When $\Delta_{B\cup\{a_i\}}(x_i, x_j) > \delta$, we get $x_i \notin \delta_{B\cup\{a_i\}}(x_j)$. In other words, x_i is not necessary to be placed in $\delta_B(x_j)$.

(2) When $\Delta_{B\cup\{a_i\}}(x_i, x_j) \leq \delta$, we get $x_i \in \delta_{B\cup\{a_i\}}(x_j)$. Suppose $\delta_{B\cup\{a_i\}}(x_j) = \{x_1, \ldots, x_i, \ldots, x_j, \ldots, x_k\}$, if we want to judge whether x_j is the positive region sample, we need to determine whether all samples in $\delta_{B\cup\{a_i\}}(x_j)$ have the same label. Since E-

q. (10), we only need to compare the labels of other samples except x_i. In other words, whether x_j is a positive region sample can not be affected by the positive region sample x_i. Thus, we also can delete x_i from the neighborhood search range $\delta_B(x_j)$.

To sum up, we can delete x_i from $\delta_B(x_j)$. The proof is completed.

For easy understanding, we give Example 1.

Example 1: Given a decision system $DS = \langle U, C, D \rangle$, where $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, \ldots, x_{10}\}$, $C = \{a_1, a_2, a_3, a_4\}$, $D = \{[1,1,1,1,1,0,0,0,0,0]^T\}$. Suppose the neighborhood of each sample is as follows.

$$\delta_{\{a_1\}}(x_1) = \{x_1, x_2, x_3, x_4, x_7\}, \tag{11}$$

$$\delta_{\{a_1\}}(x_2) = \{x_1, x_2, x_7, x_8, x_9\}, \tag{12}$$

$$\delta_{\{a_1\}}(x_3) = \{x_1, x_3, x_4, x_5\}, \tag{13}$$

$$\delta_{\{a_1\}}(x_4) = \{x_1, x_3, x_4\}, \tag{14}$$

$$\delta_{\{a_1\}}(x_5) = \{x_3, x_5, x_6, x_7\}, \tag{15}$$

$$\delta_{\{a_1\}}(x_6) = \{x_5, x_6, x_8, x_9, x_{10}\}, \tag{16}$$

$$\delta_{\{a_1\}}(x_7) = \{x_1, x_5, x_7, x_8, x_9, x_{10}\}, \tag{17}$$

$$\delta_{\{a_1\}}(x_8) = \{x_2, x_6, x_7, x_8\}, \tag{18}$$

$$\delta_{\{a_1\}}(x_9) = \{x_2, x_6, x_7, x_9\}, \tag{19}$$

$$\delta_{\{a_1\}}(x_{10}) = \{x_6, x_7, x_{10}\}. \tag{20}$$

From Definition 5 and Eqs. (11) to (20), the positive region of $B = \{a_1\}$ to D is

$$POS(D)_{\{a_1\}} = \{x_3, x_4, x_{10}\}. \tag{21}$$

Suppose $POS(D)_{\{a_1\}}$ is the biggest one in $POS(D)_{\{a_i\}}$. Thus, a_1 is selected into B. Eqs. (11) to (20) are the neighborhood search range of $\delta_{\{a_1 \cup a_i\}}(x_j)(i \neq 1)$, but there are positive region samples, i.e., x_3, x_4 and x_{10}, in them. Therefore, we need delete the positive region samples from Eqs. (11) to (20). $\delta_B(x_i)(B = \{a_1\}, x_i \notin POS(D)_{\{a_1\}})$ is as follows.

$$\delta_B(x_1) = \{x_1, x_2, x_7\}, \tag{22}$$

$$\delta_B(x_2) = \{x_1, x_2, x_7, x_8, x_9\}, \tag{23}$$

$$\delta_B(x_5) = \{x_5, x_6, x_7\}, \tag{24}$$

$$\delta_B(x_6) = \{x_5, x_6, x_8, x_9\}, \tag{25}$$

$$\delta_B(x_7) = \{x_1, x_5, x_7, x_8, x_9\}, \tag{26}$$

$$\delta_B(x_8) = \{x_2, x_6, x_7, x_8\}, \tag{27}$$

$$\delta_B(x_9) = \{x_2, x_6, x_7, x_9\}. \tag{28}$$

Eqs. (22) to (28) are the neighborhood search range of $\delta_{B \cup a_i}(x_j)(a_i \notin B)$, where x_j is the boundary region sample. In FNC, take Eq. (11) for example, we need to calculate the distance between x_1 and x_2, x_3, x_4, x_7; while in FARforBD, we only need to calculate the distance between x_1 and x_2, x_7 as in Eq. (22). In big datasets, any neighborhood may have a lot of samples. As a result, deleting the positive region sample can avoid many useless calculations with them. This can further accelerate the neighborhood search.

The algorithm of the fast attribute reduction for big datasets based on NRS is as shown in Algorithm 1. Steps 3 to 19 are generating the initial neighborhood and positive region sample judgment. In Steps 11 to 12, when the labels of x_i and x_i are different, the neighborhood search terminated in advance no matter how many samples are left behind. Step 24 is to delete the positive region sample according to the biggest $POS(D)_{B\cup\{a_i\}}$.

Algorithm 1 Fast attribute reduction for big datasets based on NRS

Input: Decision system $DS = \langle U, C, D\rangle$, δ.
Output: B.

1: $B = \emptyset$;
2: **repeat**
3:   **for** each $a_i \in (C - B)$ **do**
4:     $DS_i = \langle U, B \cup \{a_i\}, D\rangle$;
5:     **if** $B = \emptyset$ **then**
6:       $\delta_B(x_j) = U$;
7:     **end if**
8:     $\delta_{B\bigcup\{a_i\}}(x_j) = \delta_B(x_j)$;
9:     **for** each $x_k \in \delta_{B\bigcup\{a_i\}}(x_j)$ **do**
10:       Calculate $\Delta_{B\bigcup\{a_i\}}(x_j, x_k)$;
11:       **if** $\Delta_{B\bigcup\{a_i\}}(x_j, x_k) < \delta$ and the label of x_j and x_k is different **then**
12:         Break;
13:       **end if**
14:       **if** $\Delta_{B\bigcup\{a_i\}}(x_j, x_k) > \delta$ **then**
15:         Delete x_k from $\delta_{B\bigcup\{a_i\}}(x_j)$;
16:       **end if**
17:     **end for**
18:     $POS(D)_{B\cup\{a_i\}} = \emptyset$;
19:     **if** the decision attribute values in $\delta_{B\bigcup\{a_i\}}(x_j)$ are the same **then**
20:       $POS(D)_{B\cup\{a_i\}} = POS(D)_{B\cup\{a_i\}} \cup \{x_j\}$;
21:     **end if**
22:   **end for**
23:   Find the biggest $POS(D)_{B\cup\{a_i\}}$ and the corresponding a_i;
24:   **if** $POS(D)_{B\cup\{a_i\}} > 0$ **then**
25:     $B = B \cup \{a_i\}$;
26:     $\delta_B(x_j) = \delta_{B\bigcup\{a_i\}}(x_j)$;
27:     Delete the positive region sample in $\delta_B(x_j)$;
28:   **end if**
29: **until** $POS(D)_{B\cup\{a_i\}} = 0$ or $C - B = \emptyset$

4. Experiment

In order to prove the feasibility and effectiveness of FARforBD, we compare FARforBD with FNC on real datasets. In the experiments, we randomly select seven UCI benchmark datasets, whose sample sizes range from 8000 to 30000 and dimensions from 7 to 24. Each dataset is normalized. Table 1 lists the basic information of the selected UCI datasets. The experimental platform on which the algorithm runs is implemented in python 3.7 and runs in the hardware environment of Intel (R) core (TM) i7-6700 CPU @ 3.41 GHz with 32G RAM.

Table 1. Dataset Information

Datasets	Samples	Attributes	Classes
Mushroom	8124	23	2
Online shoppers	12330	18	2
Magic	19020	11	2
Letter	20000	17	26
Occupancy	20560	7	2
Avila	20867	10	12
Default	30000	24	2

In NRS, δ is an important parameter. There are many scholars have done a lot of researches on how to choose the optimal δ. Especially Hu et al. have made many experiments and verified that the optimal neighborhood parameters of different datasets are different. They also get the conclusion that the ideal value of the optimal neighborhood parameter is in [0,0.4]. There are few attributes can be found when δ is too big or too small in the NRS-based attribute reduction. However, this study is not for the optimal δ but for accelerating the neighborhood search in NRS-based attribute reduction on big datasets. Therefore, δ is set as suggested in Ref. [25], i.e., $\delta = 0.125$, to test the attribute reduction result and efficiency of FARforBD and FNC. Table 2 shows the selected attributes obtained by using FARforBD and FNC on seven big datasets. The second and third columns represent the selected attributes obtained by using the two methods, respectively. The last column is the number of attributes in the selected attributes. We can conclude that when δ is the same, the selected attributes obtained by using these two algorithms are exactly the same.

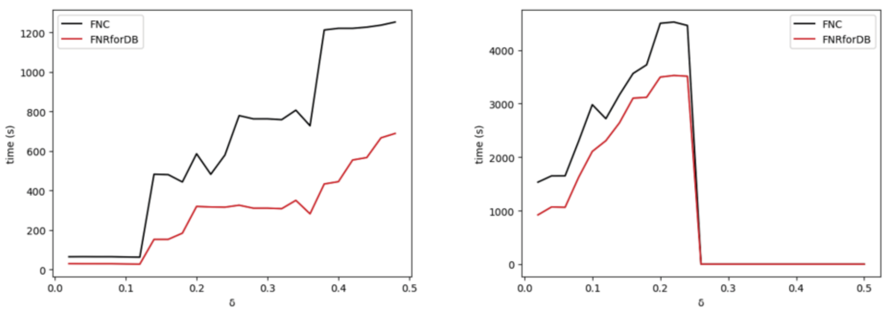
Table 2. Reduction results of FARforBD and FNC on each datasets.

Datasets	FARforBD	FNC	Number
Mushroom	$[18,4,21,10]$	$[18,4,21,10]$	4
Online shoppers	$[8,9,...,7,0,2,4]$	$[8,9,...,7,0,2,4]$	17
Magic	$[5,0,7,8,1,9,2,4,6,3]$	$[5,0,7,8,1,9,2,4,6,3]$	10
Letter	$[9,10,8,...,15,2,5,4,13,0,3]$	$[9,10,8,...,15,2,5,4,13,0,3]$	16
Occupancy	$[2,0,1,5,3,4]$	$[2,0,1,5,3,4]$	6
Avila	$[0,6,2,3,9,8,1,7,4,5]$	$[0,6,2,3,9,8,1,7,4,5]$	10
Default	$[2,6,...,20,17,16,13,19,18]$	$[2,6,...,20,17,16,13,19,18]$	23

When $\delta = 0.125$, the running time of FARforBD and FNC on each dataset is as shown in Table 3, where "Time" equals to $\frac{The\ running\ time\ of\ FNC}{The\ running\ time\ of\ FARforBD}$. The experiment results show that the running time of FARforBD on all datasets is much shorter than that of FNC. Take Mushroom for example, FARforBD only needs 294.92s, while FNC needs more than 925s. The efficiency of FARforBD is 3.13 times than that of FNC. The highest efficiency improvement is on Occupancy, because the efficiency of FARforBD is 6.18 times than that of FNC. We can conclude that FARforBD is more efficient than FNC on big datasets. To further illustrate the efficiency on different δ, we randomly select Mushroom and Letter. Figure 1 shows the running time of FARforBD and FNC on Mushroom and Letter. δ is set from 0.02 to 0.5 with a step size of 0.02. We can see that even for different δ, the efficiency of FARforBD is always better than that of FNC.

Table 3. Running time of FARforBD and FNC on each datasets.

Datasets	FARforBD(s)	FNC(s)	Times	Saved time(s)
Mushroom	294.92	925.15	3.13	630.23
Online shoppers	1397.79	2113.35	1.51	715.56
Magic	6712.81	7736.13	1.15	1023.32
Letter	3354.23	4535.01	1.35	1180.78
Occupancy	2778.68	17176.59	6.18	14897.91
Avila	3947.57	4482.48	1.14	534.91
Default	2718.96	5440.42	2.00	2721.46



(a) Mushroom (b) Letter

Figure 1. Running time comparison between FARforBD and FNC.

5. Conclusion

To improve the efficiency of NRS, narrowing the neighborhood search range down is an important factor in neighborhood calculation. Some studies have achieved good result-s in all small datasets. However, they do not perform well on big datasets. Peng et al. propose FNC, which reduces the neighborhood search range from the whole universe or the boundary region to the neighborhood of the child attribute set, and apply it to neighborhood attribute reduction. FNC has greatly improved the efficiency of neighborhood search, but it is still inefficient on big datasets. To solve this problem, we propose a fast attribute reduction method especially for big datasets based on NRS. In FARforBD, we further reduce the neighborhood search range and many useless calculations by deleting the positive region samples from the neighborhood of the child attribute set. In addition, we use many experiments on big datasets to verify the effectiveness of the proposed method. Compared to FNC, the running time of FARforBD on seven big datasets can save more than 500 seconds to 14897 seconds on the premise that the reduction results of the two algorithms are the same. The experiential results illustrate that the proposed method in this study is more efficient than the state-of-the-art comparison algorithm. In the future, we will continue to improve the efficiency of neighborhood search both in small and big datasets.

6. Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 62176033 and 61936001, the Natural Science Foundation of Chongqing under Grant No. cstc2019jcyj-cxttX0002 and by NICE: NRT for Integrated Computational Entomology, US NSF award 1631776, the National Key Research and Development Program of China under Grant No. 2019QY(Y)0301.

References

[1] Z. Pawlak, "Rough sets," International journal of computer and information sciences, vol. 11, no. 5, pp. 341–356, 1982.

[2] Z. Pawlak, "Rough set theory and its applications to data analysis," Cybernetics and Systems, vol. 29, no. 7, pp. 661–688, 1998.

[3] Y. Yang, D. Chen, and H. Wang, "Active sample selection based incremental algorithm for attribute reduction with rough sets," IEEE Transactions on Fuzzy Systems, vol. 25, no. 4, pp. 825–838, 2017.

[4] W. Ding, W. Pedrycz, I. Triguero, Z. Cao, and C. Lin, "Multigranulation supertrust model for attribute reduction," IEEE Transactions on Fuzzy Systems, vol. 29, no. 6, pp. 1395–1408, 2021.

[5] S. Salesi, G. Cosma, and M. Mavrovouniotis, "TAGA: Tabu asexual genetic algorithm embedded in a filter/filter feature selection approach for high-dimensional data," Information Sciences, vol. 565, p-p. 105–127, 2021.

[6] X. H, C. HZ, and N. DX, "Discretization of continuous attributes in rough set theory based on information entropy," Chinese Journal of Computers, vol. 28, no. 9, pp. 1570–1574, 2005.

[7] J. R and S. Q, "Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches," IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 12, pp. 1457–1471, 2004.

[8] J. Li, X. Yang, X. Song, J. Li, P. Wang, and D. Yu, "Neighborhood attribute reduction: a multi-criterion approach," International Journal of Machine Learning and Cybernetics, vol. 10, no. 4, pp. 731–742, 2019.

[9] C. Wang, Q. Hu, X. Wang, D. Chen, Y. Qian, and Z. Dong, "Feature selection based on neighborhood discrimination index," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 7, pp. 2986–2999, 2018.

[10] W. Li, Z. Huang, X. Jia, and X. Cai, "Neighborhood based decision-theoretic rough set models," International Journal of Approximate Reasoning, vol. 69, pp. 1–17, 2016.

[11] S. Xia, H. Zhang, W. Li, G. Wang, E. Giem, and Z. Chen, "GBNRS: A novel rough set algorithm for fast adaptive attribute reduction in classification," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 3, pp. 1231–1242, 2022.

[12] L. Sun, T. Wang, W. Ding, J. Xu, and Y. Lin, "Feature selection using fisher score and multilabel neighborhood rough sets for multilabel classification," Information Sciences, vol. 578, pp. 887–912, 2021.
[13] R. A. M. and E. E. Kerre, "A comparative study of fuzzy rough sets," Fuzzy Sets and Systems, vol. 126, no. 2, pp. 137–155, 2002.
[14] C. Wang, M. Shao, Q. He, Y. Qian, and Y. Qi, "Feature subset selection based on fuzzy neighborhood rough sets," Knowledge-Based Systems, vol. 111, pp. 173–179, 2016.
[15] E. C. Tsang, Q. Hu, and D. Chen, "Feature and instance reduction for pnn classifiers based on fuzzy rough sets," International Journal of Machine Learning and Cybernetics, vol. 7, pp. 1–11, 2016.
[16] C. Wang, Y. Huang, W. Ding, and Z. Cao, "Attribute reduction with fuzzy rough self-information measures," Information Sciences, vol. 549, no. 12, pp. 68–86, 2021.
[17] T. Yang, X. Zhong, G. Lang, Y. Qian, , and J. Dai, "Granular matrix: A new approach for granular structure reduction and redundancy evaluation," IEEE TRANSACTIONS ON FUZZY SYSTEMS, vol. 28, no. 12, pp. 1–1, 2020.
[18] H. Jiang, J. Zhan, and D. Chen, "Covering-based variable precision (I,T)-fuzzy rough sets with applications to multiattribute decision-making," IEEE Transactions on Fuzzy Systems, vol. 27, no. 8, pp. 1558–1572, 2019.
[19] Z. Huang and J. Li, "Discernibility measures for fuzzy covering and their application," IEEE Transactions on Cybernetics, 2021. doi:10.1109/TCYB.2021.3054742.
[20] W. Zhu and F. Wang, "Properties of the third type of covering-based rough sets," in International Conference on Machine Learning and Cybernetics, vol. 7, pp. 3746–3751, 2007.
[21] Y. Du, Q. Hu, P. Zhu, and P. Ma, "Rule learning for classification based on neighborhood covering reduction," Information Sciences, vol. 181, no. 24, pp. 5457–5467, 2011.
[22] X. Zhang, J. Wang, J. Zhan, and J. Dai, "Fuzzy measures and choquet integrals based on fuzzy covering rough sets," IEEE Transactions on Fuzzy Systems, 2021. doi:10.1109/TFUZZ.2021.3081916.
[23] Y. Tian, Z. Li, and X. Yang, "A granular reduction algorithm based on covering rough sets," Journal of Applied Mathematics, vol. 2012, no. 8, pp. 701–708, 2012.
[24] T. Yang, X. Zhong, G. Lang, Y. Qian, and J. Dai, "Granular Matrix: A new approach for granular structure reduction and redundancy evaluation," IEEE Transactions on Fuzzy Systems, vol. 28, no. 12, pp. 3133–3144, 2020.
[25] Q. Hu, D. Yu, and Z. Xie, "Numerical attribute reduction based on neighborhood granulation and rough approximation," Journal of Software, vol. 19, no. 3, pp. 640–649, 2008.
[26] Q. Hu, J. Liu, and D. Yu, "Mixed feature selection based on granulation and approximation," Knowledge-Based Systems, vol. 21, no. 4, pp. 294–304, 2008.
[27] Q. Hu, W. Pedrycz, D. Yu, and J. Lang, "Selecting discrete and continuous features based on neighborhood decision error minimization," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 40, no. 1, pp. 137–150, 2010.
[28] L. Sun, T. Yin, W. Ding, Y. Qian, and J. Xu, "Feature selection with missing labels using multilabel fuzzy neighborhood rough sets and maximum relevance minimum redundancy," IEEE Transactions on Fuzzy Systems, vol. 30, no. 5, pp. 1197–1211, 2022.
[29] B. Sang, H. Chen, L. Yang, T. Li, W. Xu, and C. Luo, "Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set," Knowledge-Based Systems, vol. 227, p. 107223, 2021.
[30] Q. Hu, H. Zhao, and D. Yu, "Efficient symbolic and numerical attribute reduction with neighborhood rough sets," Pattern Recognition and Artificial Intelligence, vol. 21, no. 6, pp. 732–738, 2008.
[31] Y. Liu, W. Huang, Y. Jiang, and Z. Zeng, "Quick attribute reduct algorithm for neighborhood rough set model," Information Sciences, vol. 271, pp. 65–81, 2014.
[32] X. Peng, P. Wang, S. Xia, C. Wang, C. Pu, and J. Qian, "FNC: a fast neighborhood calculation framework," Knowledge-Based Systems, vol. 252, p. 109394, 2022.