Fuzzy Rough Incremental Attribute Reduction Applying Dependency Measures

Yangming Liu [1] Suyun Zhao [*1,2] Hong Chen [1] and Cuiping Li [1] Yanmin Lu [3]

[1] School of Information, Renmin University of China,100872 Beijing, China
[2] Key Laboratory of Data Engineering and Knowledge Engineering (Renmin University of China), MOE, Beijing, China
[3] Building 6，1388 Zhangdong Road Pudong New Area，Shanghai，P.R.China，201203
*Corresponding Author: zhaosuyun@ruc.edu.cn

Abstract. Since data increases with time and space, many incremental rough based reduction techniques have been proposed. In these techniques, some focus on knowledge representation on the increasing data, some focus on inducing rules from the increasing data. Whereas there is less work on incremental feature selection (i.e., attribute reduction) from the increasing data, especially the increasing real valued data. And fuzzy rough sets is then applied in this incremental method because fuzzy rough set can effectively reduce attributes from the real valued data. By analyzing the basic concepts, such as lower approximation and positive region, of fuzzy rough sets on incremental datasets, the incremental mechanisms of these concepts are then proposed. An incremental algorithm is then designed. Finally, some numerical experiments demonstrate that the incremental algorithm is effective and efficient compared to non-incremental attribute reduction algorithms, especially on the datasets with large number of attributes.

Keywords: Feature selection; Incremental learning; Fuzzy rough set; Dependency function

1 Introduction

Recently, rough theory has already attracted much attention. And then many techniques based on rough set theory and its generalizations have been developed [1-4]. Most of these approaches, however, are proposed based on the assumption that the data are static. Once the data are updated or dynamically increase with time or space, these techniques have to be re-computed on the updated database, which are very costly or even intractable [5]. Incremental learning is a promising approach to refreshing data mining results, because it utilizes previously saved results or data structures to avoid the expense of re-computation [6-8].To deal with a dynamically increasing dataset, there already exist a lot of rough based researches in an incremental manner and they have successfully been used to data analysis in the real time applications and the applications with limited memory and computability [4-13].

Considering their applications, the rough based incremental approaches are split into the incremental approaches for knowledge representation, feature selection and rule induction. In rough philosophy, feature selection is also called attribute reduction. In the case of attribute reduction, some researchers constructed an incremental attribute reduction algorithm when new objects are added into a decision information system one by one [13]. Furthermore, some researchers proposed some incremental ways when objects are updated in groups [12], some other researchers conducted attribute reduction methods when an attribute set are updated [14].

By the above analysis, it is easy to see that incremental attribute reduction approaches received relatively less attention. The key reason is that it is hard to design an incremental mechanism for attribute reduction algorithms because the relation of granularity in the updated knowledge space is hard to measure before and after attribute reduction, especially for the fuzzy granularity.

This motivates the investigation of incremental attribute reduction in dynamic real valued data. Considering fuzzy rough set technique is a useful tool to handle real values, this paper develops a fuzzy rough based incremental attribute reduction algorithm in real valued dynamic datasets. Here we focus on the dynamic datasets with increasing objects, in the near future we would work on dynamic system with increasing attributes. Our main contributions in this paper include:

1) Incremental mechanisms for fuzzy positive region and dependency function are proposed by analyzing their differences before and after some new objects are added.

2) What is more, the incremental mechanism for reduction is proposed based on strict theoretical reasoning.

3)Then, we propose an incremental attribute reduction algorithm, which is effective and efficient, especially on the datasets with a large number of attributes.

The remainder of this paper is organized as follows. In Section 2, we briefly review FRS. Section 3 proposes some incremental mechanisms. Section 4 designs an incremental attribute reduction algorithm. And then in Section 5, we give some numerical experiments. Section 6 concludes this paper.

2 Preliminaries

2.1 Some Notations of Fuzzy Rough Set

The data set can be described as one decision table, denoted by a triple $DT = < U, A, D >$. Assumed that B(B ⊆ A) is one subset attributes.

Definition 1. Based on the subset B, the similarity of x_i and x_j is defined as $r_B(x_i, x_j)$, the distance of x_i and x_j is defined as $d_B(x_i, x_j)$.

(1) $d_B(x_i, x_j) = max\left\{\left|x_{B_i} - x_{B_j}\right|\right\}, 0 < i, j < |B|;$

(2) $r_B(x_i, x_j = 1 - d_B(x_i, x_j)$.

Proposition 1. In Definition 3, the positive region can be simplified as follows:

$$POS_B^U(x) = \min_{\{u \in U, u \notin D_x\}} \{d_B(x, u)\}, x \in D_x$$

Proposition 1 shows the relation between fuzzy positive region and lower approximation. Then, the dependency degree of subset B is defined as follows.

Definition 2. The dependency degree of B ⊆ A corresponding to the decision attributes D is as follows.

$$\gamma_B^U = \frac{\sum_{x\in U} POS_B^U(x)}{|U|}$$

Theorem 1. If $B_1 \subseteq B_2 \subseteq A$, then $\gamma_{B_1}^U \le \gamma_{B_2}^U \le \gamma_A^U$.

Definition 3. Given a fuzzy decision table $DT =< U, A, D >$ and attribute subset B $\subseteq$ A, when B satisfies the following two conditions, we say that B is a reduction,

(1) $\forall a \in B, \gamma_{B-a}^U < \gamma_B^U$; (2) $\gamma_B^U = \gamma_A^U$;

Then the non-incremental attribute reduction algorithm, shortened by *NonIAR*, is described in Algorithm 1 [1,2].

Algorithm 1: NonIAR
INPUT: $DT =< U, A, D >$
OUTPUT : *redu*

$B=\phi$
Lef=A
Calculate $\boldsymbol{\gamma_A^U}$
While $\boldsymbol{\gamma_B^U < \gamma_A^U}$ ***do***
    $\gamma_{B+a}^U = max\{\gamma_{B+a_i}^U\}, a_i \in lef$
    $B = B \cup a$
End while
$i \leftarrow 0$
While $i < |B|$ *do*
    if $\gamma_{B-B_i}^U < \gamma_A^U$ ***then*** $red = red \cup B_i$;
    end if
    $i{+}{+}$;
End while
Output *redu*

3 Fuzzy Rough based incremental mechanisms

3.1 Incremental Mechanism for Positive Region

Theorem 2. Given a fuzzy decision table $DT =< U, A, D >$ and an attribute subset $B \subseteq A$, If some new objects $\Delta U = \{x_{n+1}, x_{n+2}, \dots, x_{n+s}\}$ are added, then

If $x \in U$, then $POS_B^{U\cup\Delta U} = POS_B^U(x) - \Delta POS_B^U(x)$; $\Delta POS_B^U(x) = POS_B^U(x) - \min_{u\in\Delta U, U\notin D_x} d_B(x,u)$, $POS_B^U(x) > \min_{u\in\Delta U, U\notin D_x}\{d_B(x,u)\}$; $\Delta POS_B^U(x) = 0, otherwise$.

If $x \in \Delta U$, then $POS_B^U(x) = \min_{u\in U\cup\Delta U, u\notin D_x}\{d_B(x,u)\}$.

Proof.

By Proposition 2, in the case of $x \in \Delta U$, it is apparent. In the case of $x \in U$,

$POS_B^{U\cup\Delta U}(x) = \min_{u\in U\cup\Delta U, u\notin D_x}\{d_B(x,u)\} = min\{\min_{u\in U\cup\Delta U, u\notin D_x}\{d_B(x,u)\}, \min_{u\in\Delta U, u\notin D_x}\{d_B(x,u)\}\}$

$= \min_{u\in U, u\notin D_x}\{d_B(x,u)\} - \Delta POS_B^U(x) = POS_B^U(x) - \Delta POS_B^U(x)$ ■

3.2 Incremental Mechanism for Fuzzy Dependency

Theorem 3. Given a fuzzy decision table $DT =< U, A, D >$ and an attribute subset $B \subseteq A$, If some new objects ΔU are added in, then

$$\gamma_B^{U \cup \Delta U} = \frac{|U|\gamma_B^U - \sum_{x \in U} \Delta POS_B^U(x) + \sum_{x \in \Delta U} \Delta POS_B^{U \cup \Delta U}(x)}{|U| + |\Delta U|}$$

Proof.

$$\sum_{x \in U \cup \Delta U} POS_B^U(x) = \sum_{x \in U} POS_B^{U \cup \Delta U}(x) + \sum_{x \in \Delta U} POS_B^{U \cup \Delta U}(x)$$

$$= \sum_{x \in U}\left(POS_B^U(x) - \Delta POS_B^U(x)\right) + \sum_{x \in \Delta U} POS_B^{U \cup \Delta U}(x) = \sum_{x \in U} POS_B^U(x) - \sum_{x \in U} \Delta POS_B^U(x) + \sum_{x \in U} POS_B^{U \cup \Delta U}(x)$$

$$\gamma_B^{U \cup \Delta U} = \frac{\sum_{x \in U \cup \Delta U} POS_B^{U \cup \Delta U}(x)}{|U| + |\Delta U|} = \frac{\sum_{x \in U} POS_B^U(x) - \sum_{x \in U} \Delta POS_B^U(x) + \sum_{x \in \Delta U} \Delta POS_B^{U \cup \Delta U}(x)}{|U| + |\Delta U|}$$

$$= \frac{|U|\gamma_B^U - \sum_{x \in U} \Delta POS_B^U(x) + \sum_{x \in \Delta U} POS_B^{U \cup \Delta U}(x)}{|U| + |\Delta U|} \ . \ \blacksquare$$

3.3 Incremental Mechanism for reduction

Theorem 4. Given a fuzzy decision table $DT =< U, A, D >$, and an attribute subset $B \subseteq A$, when one attribute a is added into B, let $\Delta S = \{x \in U \wedge POS_B^U(x) < POS_C^U(x)\}$, then

$$\gamma_{B+a}^U = \frac{|U|\gamma_B^U - \sum_{x \in \Delta S} POS_B^U(x) + \sum_{x \in \Delta S} POS_{B+a}^U(x)}{|U|}$$

Proof.

If $B \subseteq A = \{A_1, A_2, \ldots, A_n\}, x \in U \Rightarrow \forall x \in U, POS_B^U(x) \leq POS_A^U(x)$ .If $x \notin \Delta S$, then $POS_B^U(x) = POS_A^U(x)$, which means $POS_B^U(x)$ would not grow any more.

As a result, when adding one attribute , we just need to consider

$$\gamma_{B+a}^U = \frac{|U|\gamma_B^U - \sum_{x \in \Delta S} POS_B^U(x) + \sum_{x \in \Delta S} POS_{B+a}^U(x)}{|U|}, x \in \Delta S. \ \blacksquare$$

4 Incremental Algorithm for Reduction

Algorithm 2: ***FIAR*** (Dependency Function based Incremental Reduction Algorithm)
INPUT: $DT =< U, A, D >, \Delta U, redu, \gamma_{redu}^U, \cup_{x \in U} POS_{redu}^U, \gamma_C^U, \cup_{x \in U} POS_c^U$
OUTPUT: *newredu*

$\boldsymbol{Let}\ B = redu$;

Calculate $\gamma_B^{U \cup \Delta U}$;

Calculate $\gamma_C^{U \cup \Delta U}$

$\boldsymbol{Let}\ lef = C - redu$;

$\boldsymbol{While}\ \gamma_B^{U \cup \Delta U} < \gamma_C^{U \cup \Delta U}\ do$

$besta = \{a\ |\gamma_{besta}^{U \cup \Delta U} = max\{\gamma_{a_i}^{U}\}, a_i \in lef$;

$B = B \cup besta$;

End while

$i \leftarrow 0$

$\boldsymbol{While}\ i < |B|\ do$

$\boldsymbol{If}\ notredundancy(B_i)\ \boldsymbol{then}\ \ red = red \cup B_i$;

End if

i++;

End while

Output *red*

Both NonIAR and FIAR are heuristic. The complexity of FIAR is O(M ∗|lef||ΔS||U +ΔU|), whereas NonIAR is O(K ∗N$|U + \Delta U|^2$). When N is far larger than |U|, FIAR is significantly faster than NonIAR.

5 Numerical Experiments

In this section, we conduct some numerical experiments on a series of UCI datasets [15]. The dataset detailed are summarized in Table 1. All the experiments have been carried out on Ubuntu release 16.0, i7-4790 CPU @ 3.60GHz with 8GB and c++..

Table 1. The description of the selected datasets

Datasets	Attribute No.	Object No.	Classes
Waveform	22	5000	3
Letter	17	20000	26
Shuttle	10	58000	2
Credit	25	30000	2
Gene12	9182	174	5
Gene14	3312	203	5

5.1 Execution Time

New data are added in one time.

In this part, every dataset is equally split into two parts. We use FIAR-2 represents FIAR run on such new data added in one time.

Table 2. The execution time when new data are added in one time

Datasets	***NonIAR*(CPU seconds)**	***FIAR*(CPU seconds)**	**Ratio (*FIAR/NonIAR*)**
Waveform	209.67	135	0.64
Letter	2521	1479	0.58
Shuttle	3380	1141	0.33
Credit	3633	1979	0.54

Gene12	4240	497	0.117
Gene14	876	55	0.062

Table 2 demonstrates that the ratios of FIAR/NonIAR are always smaller than 1. This shows that the incremental algorithm accelerates the reduction.

Table 2 also shows that FIAR works dramatically better than NonIAR on the datasets with high dimension. For example, FIAR is dramatically faster than NonIAR on Gene9, Gene12 and Gene14. This is because the time complexity of FIAR is $O(M * |lef||\Delta S||U + \Delta U|)$, whereas NonIAR is $O(K * N|U + \Delta U|^2)$. When N is much larger, it needs more loops for NonIAR's $\gamma_B^{U\cup\Delta U}$ approaching to $\gamma_C^{U\cup\Delta U}$, which means $M \ll K$. Thus, FIAR works significantly faster than NonIAR on the datasets with high dimension.

New data are added successively.

In his part, data is split into six parts. One part is seen as the old data, the other parts are seen as the successive added-in data. We use FIAR-6 represents it.

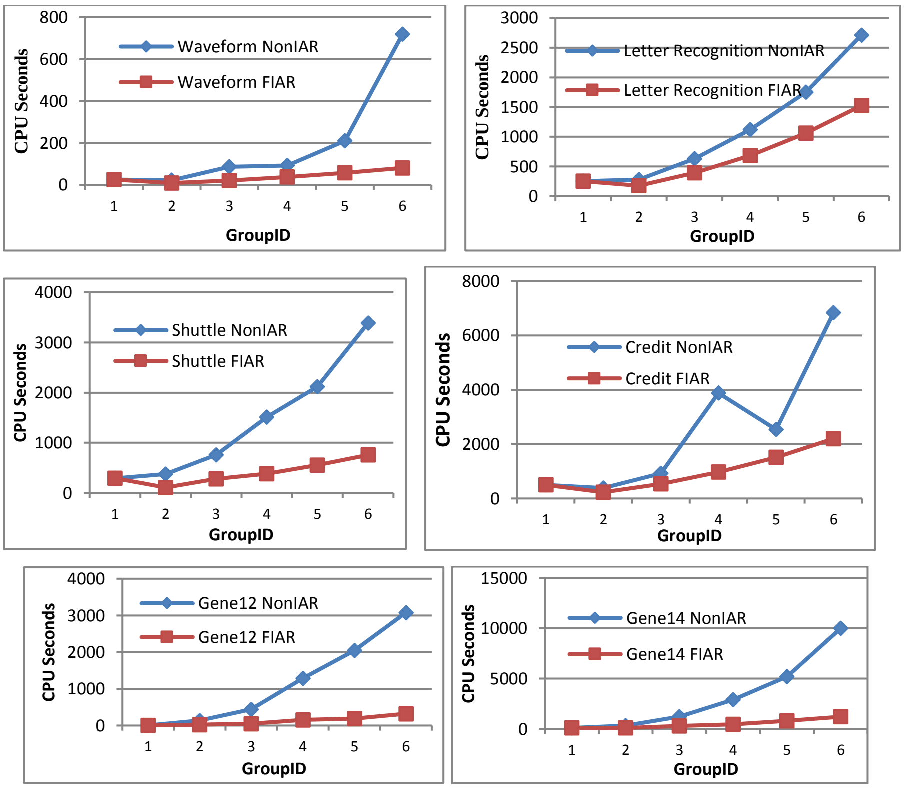

Fig. 1. The execution time on successive increasing data

In Fig 1, the trends of ***NonIAR*** go upward dramatically with the successively increasing data. Comparatively, the trends of ***FIAR*** go upward just obviously with the successively increasing data. This demonstrates that ***FIAR*** works significantly faster than ***NonIAR*** with the data successively increasing. This is because ***NonIAR*** need calculate all the data, whereas ***FIAR*** just need consider the new-added data. The faster the data dynamically update, the better FIAR works than NonIAR.

5.2 Reduction Ratio

In this part, we compare the effectiveness of reduction between ***NonIAR***, ***FIAR-2*** and ***FIAR-6***. The comparison results are summarized in Table 3.

Table 3. Reduction ratio

(a) Data sets with low dimension					(b) Dataset with high dimension				
Datasets	***NonIAR***	***FIAR-2***	***FIAR-6***	***Baseline***	***Datasets***	***NonIAR***	***FIAR-2***	***FIAR-6***	***Baseline***
Waveform	14	14	13	21					
Letter	10	9	9	17	Gene12	38	38	39	9182
Shuttle	4	5	5	10	Gene14	20	20	21	3312
Credit	8	11	10	25					
Ratio	0.49	0.53	0.50	---	Ratio	0.004	0.005	0.0049	---

In Table 3, the reduction ratios of these three algorithms are similar. They are 0.49, 0.53, 0.50 on the dataset with high dimension, and 0.004, 0.005, 0.0049 on the datasets with high dimension. This shows that the proposed incremental algorithm FIAR has the similar reduction results with the non-incremental algorithm NonIAR. Or say, the proposed incremental algorithm FIAR is an effective attribute reduction algorithm on dynamic increasing data.

5.3 Classification Performance

In this part, we apply KNN to check the classification quality of the reductions obtained by different algorithms. The comparison results are summarized in Table 4.

Table 4. The classification performance of the reductions

Datasets	NonIAR	FIAR-2	FIAR-6	Original
Waveform	66.56	66.56	66.49	67.71
Letter	77.44	77.34	77.34	77.89
Shuttle	29.16	30.02	29.56	30.82
Credit	27.68	27.13	27.71	32.94
Gene12	80.34	80.26	80.71	81.85
Gene14	70.11	70.01	70.55	72.91

In Table 4, it is easy to see that these algorithms have the similar classification accuracy. Comparing to the Baseline, i.e., the whole datasets without reduction, the accuracy lost is limited.

Table 4 also shows that the reductions of FIAR-2 and FIAR-6 have the similar classification performance. This shows that it does not affect the classification performance of reduction no matter the new data are added in one time or successively.

6 Conclusion

In this paper, we provides new views on dealing with attribute reduction on real valued datasets with the following characters:

1) Incremental mechanisms of some notations of fuzzy rough sets are proposed.

2) The incremental algorithm is designed based on fuzzy rough sets on the real valued datasets,.

3) The numerical experiments demonstrate that the proposed incremental algorithm is effective and efficient, especially on the datasets with a large number of attributes.

References

[1] Z. Pawlak. Rough sets: Theoretical aspects of reasoning about data, volume 9. Springer Science & Business Media, (2012).
[2] Z. Pawlak, A. Skowron. Rough sets: some extensions. Information sciences, 177(1):28–40, (2007).
[3] I. Düntsch, G. Gediga. Uncertainty measures of rough set prediction. Artificial intelligence, 106(1):109–137, (1998).
[4] Y. Qian, J. Liang, W. Pedrycz, C. Dang. Positive approximation: an accelerator for attribute reduction in rough set theory. Artificial Intelligence, 174(9-10):597–618, (2010).
[5] H. Chen, T. Li, C. Luo, S. Horng, G. Wang. A decision-theoretic rough set approach for dynamic data mining. IEEE Transactions on Fuzzy Systems, 23(6):1958–1970, (2015).
[6] I. Guyon, A. Elisseeff. An introduction to variable and feature selection. Journal of machine learning research, 3(Mar):1157–1182, (2003).
[7] J. Ye, Q. Li, H. Xiong, H. Park, R.Janardan, V. Kumar. IDR/QR: An incremental dimension reduction algorithm via qr decomposition. IEEE Transactions on Knowledge and Data Engineering, 17(9):1208–1222, (2005).
[8] Y. Zhang, S. Chen, Q. Wang, G. Yu. mapreduce: Incremental mapreduce for mining evolving big data. IEEE transactions on knowledge and data engineering, 27(7):1906–1919, (2015).
[9] J. Zhang, T. Li, D. Ruan, D. Liu. Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems. International Journal of Approximate Reasoning, 53(4):620–635, (2012).
[10] A. Zeng, T. Li, J. Hu, H. Chen, C. Luo. Dynamical updating fuzzy rough approximations for hybrid data under the variation of attribute values. Information Sciences, 378:363–388, (2017).
[11] J. Zhang, J. Wong, Y. Pan, T. Li. A parallel matrix-based method for computing approximations in incomplete information systems. IEEE Transactions on Knowledge and Data Engineering, 27(2):326–339, (2015).
[12] J. Liang, F. Wang, C. Dang, Y. Qian. A group incremental approach to feature selection applying rough set technique. IEEE Transactions on Knowledge and Data Engineering, 26(2):294–308, (2014).
[13] F. Hu, G. Wang, H. Huang, Y. Wu. Incremental attribute reduction based on elementary sets. In International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, pages 185–193. Springer, (2005).
[14] A. Zeng, T. Li, D.Liu, J. Zhang, H. Chen. A fuzzy rough set approach for incremental feature selection on hybrid information systems. Fuzzy Sets and Systems, 258:39–60, (2015).
[15] https://archive.ics.uci.edu/ml/datasets.html/.