LETTER

Novel Improvements on the Fuzzy-Rough QuickReduct Algorithm

Javad Rahimipour ANARAKI[†a)], **Mahdi EFTEKHARI**[††b)], *Nonmembers*, *and* **Chang Wook AHN**[†c)], *Member*

SUMMARY Feature Selection (FS) is widely used to resolve the problem of selecting a subset of information-rich features; Fuzzy-Rough QuickReduct (FRQR) is one of the most successful FS methods. This paper presents two variants of the FRQR algorithm in order to improve its performance: 1) Combining Fuzzy-Rough Dependency Degree with Correlation-based FS merit to deal with a dilemma situation in feature subset selection and 2) Hybridizing the newly proposed method with the threshold based FRQR. The effectiveness of the proposed approaches are proven over sixteen UCI datasets; smaller subsets of features and higher classification accuracies are achieved.

key words: *fuzzy-rough set, dependency degree, feature selection, fuzzy-rough quickreduct*

1. Introduction

Many problems in machine learning and pattern recognition involve high-dimensional datasets. This would be even worse when the number of features is greater than the number of objects in the dataset, due to the problem of 'curse of dimensionality'. This characteristic makes further processing infeasible and impossible. One solution to this problem is Feature Selection (FS) which tries to select information-rich features out of redundant and irrelevant features. Redundant features are the ones which have the same information as others, while irrelevant features are those which do not contain any information about outcome. It is natural to eliminate these two kinds of features so as to minimize the number of features to those which have more impact on the classification result [1].

Some of the FS methods (e.g., transformation based methods) destroy semantics of data while others (e.g., statistical correlation-based approaches) need additional information. This implies that an approach which preserves semantics of data and needs no human provided information is desired. Rough Set theory is a promising alternative to the above requirements. Selecting m features out of N ones by means of comprehensive search is an NP-hard problem as $N!/[m!(N-m)!]$. Even worse, it can be proven that approximating the minimal relevant subset of features is hard up to a very large factor [2]; thus, greedy search methods are suitable to overcome this problem.

A novel Fuzzy-Rough FS (FRFS) technique which is guided by fuzzy entropy was proposed by Parthalain *et al.* [3]. They proposed another approach which uses the information gathered from both lower approximation dependency value and the number of objects in the boundary region and the distance of those objects from the lower approximation [4]. Chen *et al.* introduced a new Rough Set approach [5] to FS based on Ant Colony Optimization (ACO), which adopts mutual information based feature significance. A novel heuristic algorithm was developed by employing the appearing frequency of attribute as heuristic information [6].

This paper enhances the conventional Fuzzy-Rough QuickReduct (FRQR) [7] by 1) combining Fuzzy-Rough Dependency Degree (FDD) with correlation based merit [1] in a dilemma situation of selecting the best feature and 2) hybridizing the proposed method with the threshold based FRQR (T-FRQR) [8].

2. Rough Set Based Feature Selection

Rough Set theory was proposed by Pawlak as a tool to deal with uncertainty. Let $\mathbb{U}$ be the universe of discourse and R be the equivalence relation on $\mathbb{U}$; approximation space is represented by $(\mathbb{U}, R)$. Let X be a subset of $\mathbb{U}$ and P be a subset of features A; approximation of this subset using Rough Set theory is conducted by means of lower and upper approximation. Objects in lower approximation ($\underline{P}X$) are surely classified into X with regard to attributes in P. Upper approximation of X ($\overline{P}X$) contains objects which are possibly classified into X considering attributes in P. Let P and Q be subsets of A, then the dependency of Q on P is denoted by $P \Rightarrow_\kappa Q$, where κ is Dependency Degree (DD) as given by

$$\kappa = \gamma_P(Q) = \frac{|POS_P(Q)|}{|\mathbb{U}|}. \quad (1)$$

Here, the notation $|\cdot|$ is used for cardinality. Positive region of the partition $\mathbb{U}/Q$ with respect to P, denoted by $|POS_P(Q)|$, is the set of all elements which can be classified into partition $\mathbb{U}/Q$ using P [7].

3. Fuzzy-Rough Set Based Feature Selection

In general, datasets which contain both crisp and real-valued data cannot be handled by Rough Set [7]. The need for a

Manuscript received May 20, 2014.
Manuscript revised September 22, 2014.
Manuscript publicized October 21, 2014.

[†]The authors are with the Department of Computer Science and Engineering, Sungkyunkwan University, Korea.
[††]The author is with the Department of Computer Engineering, Shahid Bahonar University of Kerman, Iran.
a) E-mail: j.r.anaraki@ieee.org
b) E-mail: m.eftekhari@uk.ac.ir
c) E-mail: cwan@skku.edu

DOI: 10.1587/transinf.2014EDL8099

Algorithm 1: Fuzzy-Rough QuickReduct (FRQR)

C, the set of all conditional attributes
D, the set of decision attributes
$R \leftarrow \{\}; \gamma'_{best} = 0; \gamma'_{prev} = 0$
do
  $T \leftarrow R$
  $\gamma'_{prev} \leftarrow \gamma'_{best}$
  foreach $x \in (C - R)$
    if $\gamma'_{R\cup\{x\}}(D) > \gamma'_T(D)$
      $T \leftarrow R \cup \{x\}$
      $\gamma'_{best} \leftarrow \gamma'_T(D)$
  $R \leftarrow T$
until $\gamma'_{best} = \gamma'_{prev}$
return R

method to handle all kinds of data motivated researchers to combine Fuzzy and Rough Set theories. Definitions of Fuzzy lower and upper approximations are given in Eqs. (2) and (3).

$$\mu_{\underline{R_P}X}(x) = \inf_{y\in\mathbb{U}} I\{\mu_{R_P}(x,y), \mu_X(y)\} \tag{2}$$

$$\mu_{\overline{R_P}X}(x) = \sup_{y\in\mathbb{U}} T\{\mu_{R_P}(x,y), \mu_X(y)\} \tag{3}$$

where I is Lukasiewics Fuzzy implicator which is defined by $min(1 - x + y, 1)$ and T is Lukasiewics Fuzzy t-norm which is given by $max(x + y - 1, 0)$.

$$\mu_{R_a}(x,y) \tag{4}$$
$$= max\left\{min\left\{\frac{(a(y)-(a(x)-\sigma_a))}{(a(x)-(a(x)-\sigma_a))}, \frac{((a(x)+\sigma_a)-a(y))}{((a(x)+\sigma_a)-a(x))}\right\}, 0\right\}$$

$$\mu_{R_P}(x,y) = \bigcap_{a\in P}\{\mu_{R_a}(x,y)\} \tag{5}$$

where σ_a is the variance of feature a, R_P is Fuzzy similarity relation, and $\mu_{R_a}(x,y)$ is the degree of similarity between objects x and y considering feature a. Moreover, fuzzy positive region and FDD are defined in Eqs. (6) and (7), respectively [7].

$$\mu_{POS_{R_P}(Q)}(x) = \sup_{X\in\mathbb{U}/Q} \mu_{\underline{R_P}X}(x) \tag{6}$$

$$\gamma'_P(Q) = \frac{|\mu_{POS_{R_P}(Q)}(x)|}{|\mathbb{U}|} = \frac{\sum_{x\in\mathbb{U}} \mu_{POS_{R_P}(Q)}(x)}{|\mathbb{U}|} \tag{7}$$

Based on the FDD, Fuzzy version of QuickReduct is shown in the Algorithm 1.

4. Proposed Enhancements

By monitoring behavior of FRQR in facing different datasets, some undetermined situations are raised and remain unseen due to the nature of FRQR which is greedy forward algorithm. However, these situations can be handled to improve its performance by employing proper methods. In this section, one of these situations is found and solved by employing FS merit, and then hybridized with previously proposed T-FRQR [8]. Two enhancements on the FRQR algorithm are presented as follows:

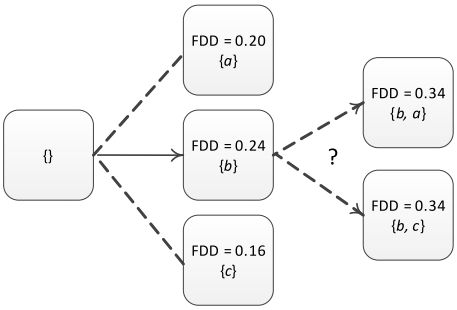

Fig. 1 An example of equal situation.

- Wise Fuzzy-Rough QuickReduct (W-FRQR)
- Hybridization of Threshold & Wise FRQR (H-FRQR)

4.1 Wise Fuzzy-Rough QuickReduct (W-FRQR)

One of the main drawbacks of the FRQR algorithm is *Equal Situation* which is depicted in Fig. 1. This situation arises when the algorithm faces more than one subset with the same dependency value. Conventional algorithm simply selects the first feature subset regarding its nature, but the other subsets might have more influence on classification accuracy. In Fig. 1, FS starts with an empty subset and continues by calculating FDD of each subset. In the first step, feature b is selected owing to its highest impact on FDD. Next, FDD of combination of feature b with remaining features (i.e., a and c) are calculated. As mentioned above, both combinations have the same increase in FDD. In this situation, traditional remedy selects the first combination with the greatest value; thus, $\{b, a\}$ is selected with no proper examination on the quality of subset $\{b, c\}$. To overcome this problem, a combination of FDD and correlation-based heuristic [1] is suggested. When FRQR confronts the *Equal Situation*, the correlation-based merit as given in Eq. (8) is calculated and a subset with the highest value is chosen.

$$Merit_S = \frac{k\overline{r}_{cf}}{\sqrt{k + k(k-1)\overline{r}_{ff}}} \tag{8}$$

where $Merit_S$ is the heuristic of selected subset S, k is the number of features in S, $\overline{r}_{cf}$ is the mean of feature-class correlation, and $\overline{r}_{ff}$ is the mean of feature-feature correlation. The proposed method is presented in Algorithm 2.

4.2 Hybridization of Threshold & Wise FRQR (H-FRQR)

In our previous work, the cardinality of selected subset was controlled using a threshold. Since adding a new feature to the reduct subset is not only time consuming while encountering dimensional datasets, but might also increase FDD too slightly; therefore, this process can be stopped by employing a threshold. This stopping criterion is formulated as follows:

$$(\gamma'_{overall} - \gamma'_{best}) \times |\mathbb{U}| < 1 \tag{9}$$

Algorithm 2: Wise FRQR (W-FRQR)

C, the set of all conditional attributes
D, the set of decision attributes
$R \leftarrow \{\}; \gamma'_{best} = 0; \gamma'_{prev} = 0;$
do
 $T \leftarrow R$
 $\gamma'_{prev} \leftarrow \gamma'_{best}$
 $X \leftarrow (C - R)$
 forall $x \in X$ **calculate** FDD_x
  if $|Max(FDD_X)| > 1$
   foreach x **calculate** $Merit_S(x)$
   $x \equiv max(Merit_S)$
  elseif $|Max(FDD_X)| = 1$
   $x \equiv max(FDD_X)$
  $T \leftarrow R \cup \{x\}$
  $\gamma'_{best} \leftarrow \gamma'_T(D)$
  $R \leftarrow T$
until $\gamma'_{best} = \gamma'_{prev}$
return R

Algorithm 3: Hybridization of Threshold & Wise FRQR (H-FRQR)

C, the set of all conditional attributes
D, the set of decision attributes
$R \leftarrow \{\}; \gamma'_{best} = 0; \gamma'_{prev} = 0; \gamma'_{overall}$
do
 $T \leftarrow R$
 $\gamma'_{prev} \leftarrow \gamma'_{best}$
 $X \leftarrow (C - R)$
 forall $x \in X$ **calculate** FDD_x
  if $|Max(FDD_X)| > 1$
   foreach x **calculate** $Merit_S(x)$
   $x \equiv max(Merit_S)$
  elseif $|Max(FDD_X)| = 1$
   $x \equiv max(FDD_X)$
  $T \leftarrow R \cup \{x\}$
  $\gamma'_{best} \leftarrow \gamma'_T(D)$
  $R \leftarrow T$
until $(\gamma'_{overall} - \gamma'_{best}) \times |\mathbb{U}| < 1$
return R

where $\gamma'_{overall}$ is the overall FDD which is calculated in the presence of all features of dataset with the complexity of $O((n^2 + n)/2)$ where n is the number of features, γ'_{best} is the current FDD of reducted subset, and $|\mathbb{U}|$ is the cardinality of dataset. A new hybridized method which benefited from both T-FRQR and W-FRQR is introduced to improve the performance of FRQR by selecting less features, and taking a wise manner in *Equal Situation*. By referring to the fundamentals of both ideas, one can expect more contribution of T-FRQR than W-FRQR in H-FRQR outcome. The H-FRQR is shown in Algorithm 3. It is obvious that the complexity of all proposed methods can be bounded by $O((n^2 + n)/2)$ where n is the number of features of dataset.

5. Experimental Results

Sixteen UCI datasets as depicted in Table 1, were employed to evaluate each methods' performance. The FRQR, T-FRQR and two proposed successors were applied to select information-rich features. The results of all methods as well as unreducted datasets in term of the number of selected features are demonstrated in Table 2.

Table 1 Datasets specifications.

No	Datasets	Instances	Attributes
1	Blood Transfusion	748	5
2	Breast Cancer	699	10
3	Breast Tissue	106	10
4	Cleveland	297	14
5	Glass	214	10
6	Heart	270	13
7	Ionosphere	351	34
8	Libras Movement	360	91
9	Lung Cancer	32	56
10	Olitos	120	26
11	Parkinson	197	23
12	Pima Indian Diabetes	768	8
13	Sonar	208	60
14	Soybean	47	35
15	SPECTF Heart	80	45
16	Wine	178	13

Table 2 Number of selected features.

Datasets	Unred.	FRQR	T-FRQR	W-FRQR	H-FRQR
Blood Transfusion	5	4	3	3	3
Breast Cancer	10	7	5	7	5
Breast Tissue	10	9	8	8	8
Cleveland	14	11	6	5	6
Glass	10	9	8	8	8
Heart	13	7	6	7	6
Ionosphere	34	7	6	7	6
Libras Movement	91	2	6	8	6
Lung Cancer	56	6	5	6	5
Olitos	26	5	4	5	4
Parkinson	23	5	4	5	4
Pima Indian Diabetes	8	8	7	7	7
Sonar	60	5	4	5	4
Soybean	35	2	1	2	1
SPECTF Heart	45	5	4	5	4
Wine	13	5	4	5	4

Nine classifiers of different categories such as PART, JRip, Naïve Bayes, Bayes Net, J48, BFTree, FT, NBTree and RBFNetwrok were selected to classify resulting subsets of features by each method. The results are presented in Table 3 where the mean of classification accuracies of nine classifiers for each dataset and each method along with unreducted datasets is shown in each cell, and the last row indicates the mean of the mean of classification accuracies. The lowest mean was gained by FRQR and ranking order to the best was H-FRQR, T-FRQR, W-FRQR and unreducted datasets. The highest mean of classification accuracies was gained in expense of employing all features, which means that feature selection methods are not always successful in increasing classification accuracy, but in decreasing model complexity by sacrificing inconsiderable accuracy. As Table 3 shows, for seven datasets the original model gained the highest classification accuracy, six for FRQR, five for T-FRQR and four for both W-FRQR and H-FRQR.

Since both the classification accuracies and the number of selected features are important, divisions of classification accuracies (Table 3) by the number of selected features (Ta-

Table 3 Mean of classification accuracies (%).

Datasets	Unred.	FRQR	T-FRQR	W-FRQR	H-FRQR
Blood Transfusion	77.20	66.46	66.25	66.26	66.25
Breast Cancer	96.18	96.23	96.49	96.29	96.49
Breast Tissue	66.46	66.46	66.25	66.25	66.25
Cleveland	50.13	49.76	51.52	51.22	50.99
Glass	61.89	67.29	64.64	64.64	64.64
Heart	79.55	78.48	73.25	78.48	73.25
Ionosphere	89.68	91.39	90.09	90.57	90.09
Libras Movement	61.70	21.76	52.10	45.31	52.10
Lung Cancer	55.56	58.85	51.44	58.85	51.44
Olitos	69.81	66.39	67.96	66.39	67.96
Parkinson	82.34	85.07	86.38	85.07	86.38
Pima Indian Diabetes	75.00	75.00	75.46	75.46	75.46
Sonar	67.47	69.82	70.94	69.82	70.94
Soybean	98.58	100.00	85.34	100.00	85.34
SPECTF Heart	73.06	64.86	65.56	70.42	65.56
Wine	85.32	95.63	94.88	94.51	94.88
Mean	**74.37**	**72.09**	**72.41**	**73.72**	**72.38**

Table 4 Division of classification accuracies and number of selected features.

Datasets	Unred.	FRQR	T-FRQR	W-FRQR	H-FRQR
Blood Transfusion	15.44	16.62	22.08	22.08	22.08
Breast Cancer	9.61	13.75	19.30	13.76	19.30
Breast Tissue	6.64	7.38	8.28	8.28	8.28
Cleveland	3.58	4.52	8.59	10.24	8.50
Glass	6.18	7.48	8.08	8.08	8.08
Heart	6.11	11.21	11.21	11.21	11.21
Ionosphere	2.63	13.06	15.02	12.94	15.02
Libras Movement	0.67	10.88	8.68	5.66	8.68
Lung Cancer	0.99	9.81	10.29	9.81	10.29
Olitos	2.68	13.28	16.99	13.28	16.99
Parkinson	3.58	17.01	21.60	17.01	21.60
Pima Indian Diabetes	9.38	9.38	10.78	10.78	10.78
Sonar	1.12	13.96	17.74	13.96	17.74
Soybean	2.81	50.00	85.34	50.00	85.34
SPECTF Heart	1.62	12.97	16.39	14.08	16.39
Wine	6.56	19.13	23.72	18.90	23.72
Mean	**4.98**	**14.40**	**19.01**	**15.00**	**19.00**

Table 5 Average rankings of the algorithms (Friedman).

Algorithm	Ranking
T-FRQR	1.7812
H-FRQR	1.8438
W-FRQR	2.9385
FRQR	3.4688
Unred.	4.9688

ble 2) were considered as a measure to compare the results. Therefore, a method which ends to the highest classification accuracy and the minimum number of features is regarded as the best method. The results are shown in Table 4 where original datasets achieved the lowest value, T-FRQR reached the highest value, and the performance of both T-FRQR and H-FRQR were nearly identical. As for the results in Table 4, a non-parametric statistical analysis was conducted to compare all the algorithms. Average ranking obtained by each method in the Friedman test are presented in Table 5. As expected, the rankings were the same as the rankings which were given in Table 4. Friedman statistic (distributed according to the chi-square with 4 degrees of freedom) is 44.3, and P-value computed by Friedman Test was 0.

6. Concluding Remarks

Two novel successors to Fuzzy-Rough QuickReduct were presented in this paper. Experimental results over sixteen datasets taken from UCI apparently showed the applicability and effectiveness of proposed methods over the conventional methods. The number of selected features, classification accuracies, and the division of classification accuracies by the number of selected features were used to measure the performance of each algorithm. The Friedman test was utilized for ranking the algorithms based on the statistical fundamentals and the aforementioned division. The T-FRQR and H-FRQR outperformed W-FRQR and FRQR in terms of the number of selected features. Although W-FRQR was placed third, but the method led to the highest classification accuracy among the proposed methods. According to the number of selected features, the classification accuracies and the statistical results, it is natural to recommend T-FRQR and H-FRQR for handling real-world applications. More investigations on the hybridization of FDD and newly emerged FS merits can be made to deal with the dilemma of selecting a subset among those with the same FDD value.

Acknowledgment

This work was supported by the NRF grant funded by the Korean government (MEST) (NRF-2012R1A2A2A01013735).

References

[1] M.A. Hall and L.A. Smith, “Feature subset selection: A correlation based filter approach,” Proc. Int. Conf. on Neural Information Processing and Intelligent Information Systems, pp.855–858, New Zealand, 1997.

[2] R. Nock and M. Sebban, “Sharper bounds for the hardness of prototype and feature selection,” in Algorithmic Learning Theory, ed. H. Arimura, S. Jain, and A. Sharma, Lecture Notes in Computer Science, vol.1968, pp.224–238, Springer, 2000.

[3] N. Parthalain, R. Jensen, and Q. Shen, “Fuzzy entropy-assisted fuzzy-rough feature selection,” IEEE Int. Conf. on Fuzzy Systems, pp.423–430, 2006.

[4] N. Parthalain, Q. Shen, and R. Jensen, “A distance measure approach to exploring the rough set boundary region for attribute reduction,” IEEE Trans. Knowl. Data Eng., vol.22, no.3, pp.305–317, March 2010.

[5] Y. Chen, D. Miao, and R. Wang, “A rough set approach to feature selection based on ant colony optimization,” Pattern Recognit. Lett., vol.31, no.3, pp.226–233, 2010.

[6] J. Zhang, J. Wang, D. Li, H. He, and J. Sun, “A new heuristic reduct algorithm base on rough sets theory,” in Advances in Web-Age Information Management, ed. G. Dong, C. Tang, and W. Wang, Lecture Notes in Computer Science, vol.2762, pp.247–253, Springer, 2003.

[7] R. Jensen and Q. Shen, “New approaches to fuzzy-rough feature selection,” IEEE Trans. Fuzzy Systems, vol.17, no.4, pp.824–838, Aug. 2009.

[8] J. Anaraki and M. Eftekhari, “Improving fuzzy-rough quick reduct for feature selection,” 19th Iranian Conf. on Electrical Engineering (ICEE), pp.1502–1506, May 2011.