Semi-Supervised
Fuzzy-Rough Feature Selection

Richard Jensen[1], Sarah Vluymans[2,3], Neil Mac Parthaláin[1], Chris Cornelis[2,4], and Yvan Saeys[3,5]

[1] Dept. of Computer Science
Aberystwyth University
Aberystwyth, Ceredigion, Wales, UK
`{rkj,ncm}@aber.ac.uk`
[2] Department of Applied Mathematics, Computer Science and Statistics
Ghent University, Belgium
`{Sarah.Vluymans,Chris.Cornelis}@ugent.be`
[3] VIB Inflammation Research Center, Zwijnaarde, Belgium
`{Sarah.Vluymans, Yvan.Saeys}@irc.vib-ugent.be`
[4] Dept. of Computer Science and AI
CITIC-UGR, University of Granada, Spain
`Chris.Cornelis@decsai.ugr.es`
[5] Department of Respiratory Medicine, Ghent University, Belgium
`Yvan.Saeys@ugent.be`

Abstract. With the continued and relentless growth in dataset sizes in recent times, feature or attribute selection has become a necessary step in tackling the resultant intractability. Indeed, as the number of dimensions increases, the number of corresponding data instances required in order to generate accurate models increases exponentially. Fuzzy-rough set-based feature selection techniques offer great flexibility when dealing with real-valued and noisy data; however, most of the current approaches focus on the supervised domain where the data object labels are known. Very little work has been carried out using fuzzy-rough sets in the areas of unsupervised or semi-supervised learning. This paper proposes a novel approach for semi-supervised fuzzy-rough feature selection where the object labels in the data may only be partially present. The approach also has the appealing property that any generated subsets are also valid (super)reducts when the whole dataset is labelled. The experimental evaluation demonstrates that the proposed approach can generate stable and valid subsets even when up to 90% of the data object labels are missing.

Keywords: Fuzzy-rough sets, feature selection, semi-supervised learning.

1 Introduction

Supervised learning operates on labelled data and attempts to learn the underlying functional relationships in that data. It is the most common paradigm in

machine learning and is concerned with the learning of classifiers which can accurately reflect the predictive regularities of the underlying model from the feature values and decision class labels. For unsupervised learning, on the other hand, there are no decision class labels and the task is to construct or reconstruct class information from some inherent structure in the data. These techniques attempt to find groups in the data such that objects in the same group are similar to each other in some way and those in different groups are dissimilar. The notion of similarity is however subjective and as such, unsupervised learning approaches are forced to make assumptions about groupings as well as the number of groups into which data objects belong. The semi-supervised learning (SSL) paradigm lies between that of supervised learning and unsupervised learning. It is typically employed when some (but not all) of the data is labelled. The primary aim of SSL is to try to utilise both labelled and unlabelled data and it has therefore attracted much interest due to the abundance of unlabelled data which is available for many real-world problems. The main obstacle for traditional learning methods is that they cannot utilise unlabelled data for knowledge discovery. This has led to a growth in the number of SSL approaches.

Rough sets [7] and fuzzy-rough sets [3] have recently enjoyed much attention particularly for the task of feature selection (FS), due to their domain independence and, in the case of fuzzy-rough sets, the additional ability to handle real-valued data. The vast majority of work carried out in the areas of rough sets and fuzzy-rough sets has been focused on supervised learning approaches, i.e. where all of the class labels are known. There has been very little work in the area of unsupervised learning for fuzzy-rough sets and even less still for semi-supervised learning. The motivation for a fuzzy-rough based semi-supervised feature selection approach is based on the success of the supervised approaches [5, 6] and the fact that the subsets produced by the proposed approaches are also shown to be valid (super)reducts for fully labelled data.

The remainder of the paper is structured as follows. In Section 2, the preliminaries for fuzzy-rough set theory and FS are covered. In Section 3, the proposed approach for semi-supervised fuzzy-rough set FS is presented. Section 4 details an experimental evaluation of the technique, where its performance is assessed through the random removal of class labels from a number of benchmark datasets and using non-parametric statistical analysis. Finally, in Section 5, the paper is concluded and some directions for future work are suggested.

2   Rough and fuzzy-rough set theory

In rough set analysis [7], data is represented as an *information system* $(X, \mathcal{A})$, where $X = \{x_1, \ldots, x_n\}$ and $\mathcal{A} = \{a_1, \ldots, a_m\}$ are finite, non-empty sets of objects and features, respectively. Each $a \in \mathcal{A}$ corresponds to a mapping from X to V_a, which is the value set of a over X. For every subset B of $\mathcal{A}$, the B-indiscernibility relation[6] R_B is defined as

$$R_B = \{(x, y) \in X^2 \mid (\forall a \in B)(a(x) = a(y))\}. \tag{1}$$

---

[6] When $B = \{a\}$, i.e., B is a singleton, R_a is written rather than $R_{\{a\}}$

Clearly, R_B is an equivalence relation. Its equivalence classes $[x]_{R_B}$ can be used to approximate concepts, i.e., subsets of the universe X. Given $A \subseteq X$, its lower and upper approximation w.r.t. R_B are respectively defined as

$$R_B{\downarrow}A \;=\; \{x \in X \mid [x]_{R_B} \subseteq A\} \tag{2}$$
$$R_B{\uparrow}A \;=\; \{x \in X \mid [x]_{R_B} \cap A \neq \emptyset\}. \tag{3}$$

An element x belongs to the lower approximation if it belongs to A and all other instances in its equivalence class do so as well. It belongs to the upper approximation when it does not necessarily belong to A itself, but there is at least one element in its equivalent class that does.

A *decision system* $(X, \mathcal{A} \cup \{d\})$ is a special kind of information system, used in the context of classification. Attribute d ($d \notin \mathcal{A}$) is called the decision feature. Its equivalence classes $[x]_{R_d}$ (or $[x]_d$) are called decision classes. Given $B \subseteq \mathcal{A}$, the B-positive region POS_B contains those objects from X for which the values of B allow to predict the decision class unequivocally. This can be modeled using the lower approximation (2), i.e.,

$$POS_B = \bigcup_{x \in X} R_B{\downarrow}[x]_{R_d}. \tag{4}$$

Indeed, if $x \in POS_B$, it means that whenever an object has the same values as x for the features in B, it will also belong to the same decision class as x. The predictive ability w.r.t. d of the features in B is measured by the following value (degree of dependency of d on B):

$$\gamma_B = \frac{|POS_B|}{|X|}. \tag{5}$$

$(X, \mathcal{A} \cup \{d\})$ is called *consistent* if $\gamma_\mathcal{A} = 1$. A subset B of $\mathcal{A}$ is called a *decision reduct* if it satisfies $POS_B = POS_\mathcal{A}$, i.e., B preserves the decision making power of $\mathcal{A}$, and moreover it cannot be further reduced, i.e., there exists no proper subset B' of B such that $POS_{B'} = POS_\mathcal{A}$. When the latter constraint is removed, i.e. B is not necessarily minimal, this is then termed a decision superreduct.

Fuzzy-rough set theory extends the above notions. A subset B of $\mathcal{A}$ can be defined using the fuzzy B-indiscernibility relation:

$$R_B(x, y) = \mathcal{T}(\underbrace{R_a(x, y)}_{a \in B}), \tag{6}$$

in which $\mathcal{T}$ represents a t-norm[7]. It can easily be seen that if only qualitative features (possibly originating from discretisation) are used, then the traditional concept of the B-indiscernibility relation is recovered. For the lower and upper

---

[7] A t-norm $\mathcal{T}$ is an increasing, commutative, associative $[0,1]^2 \to [0,1]$ mapping satisfying $\mathcal{T}(x, 1) = x$ for x in $[0, 1]$.

approximation of a fuzzy set A in X by means of a fuzzy tolerance relation R, the definitions proposed in [8] are adopted and defined, for all x in X,

$$(R{\downarrow}A)(x) = \inf_{y \in X} \mathcal{I}(R(x,y), A(y)) \tag{7}$$

$$(R{\uparrow}A)(x) = \sup_{y \in X} \mathcal{T}(R(x,y), A(y)). \tag{8}$$

Here, $\mathcal{I}$ is an implicator[8]. When d is crisp, the fuzzy positive region can be defined as follows [2]:

$$POS_B(x) = (R_B{\downarrow}[x]_d)(x). \tag{9}$$

Using the Łukasiewicz implicator $((\forall a, b \in [0,1])(\mathcal{I}(a,b) = \min(1 - a + b, 1)))$, it is found for each data instance $x \in X$:

$$
\begin{aligned}
POS_B(x) &= (R_B \downarrow [x]_d)(x) \\
&= \inf_{y \in X} \mathcal{I}(R_B(x,y), [x]_d(y)) \\
&= \min[\inf_{y \in [x]_d} \mathcal{I}(R_B(x,y), [x]_d(y)), \inf_{y \notin [x]_d} \mathcal{I}(R_B(x,y), [x]_d(y))] \\
&= \min[\inf_{y \in [x]_d} \mathcal{I}(R_B(x,y), 1), \inf_{y \notin [x]_d} \mathcal{I}(R_B(x,y), 0)] \\
&= \min[\inf_{y \in [x]_d} \min(1 - R_B(x,y) + 1, 1), \inf_{y \notin [x]_d} \min(1 - R_B(x,y) + 0, 1)] \\
&= \min[1, \inf_{y \notin [x]_d} (1 - R_B(x,y))] \\
&= \inf_{y \notin [x]_d} (1 - R_B(x,y)). \tag{10}
\end{aligned}
$$

From this, an increasing $[0,1]$-valued measure to gauge the degree of dependency of a subset of features on another subset of features can be defined. For FS, it is useful to phrase this in terms of the dependency of the decision feature on a subset of the conditional features:

$$\gamma_B = \frac{|POS_B|}{|X|} = \frac{\sum_{x \in X} POS_B(x)}{|X|}. \tag{11}$$

This measure is used in the fuzzy-rough FS method (FRFS) of [5]. This technique is a hill-climbing algorithm to determine a (super)reduct $B \subseteq \mathcal{A}$. It initialises B as an empty set and iteratively adds the attribute $a \in \mathcal{A} \setminus B$ that leads to the largest increase in the value (11). The algorithm halts when $\gamma_B = \gamma_{\mathcal{A}}$.

3 Semi-supervised Fuzzy-Rough Feature Selection

One of the primary motivating factors for semi-supervised approaches is the abundance of unlabelled data. Indeed, it is often expensive and time-consuming

---

[8] An implicator $\mathcal{I}$ is a $[0,1]^2 \to [0,1]$ mapping that is decreasing in its first and increasing in its second argument, satisfying $\mathcal{I}(0,0) = \mathcal{I}(0,1) = \mathcal{I}(1,1) = 1$ and $\mathcal{I}(1,0) = 0$.

for domain experts to label data and this is where semi-supervised techniques can take advantage of small amounts of labelled data and (larger) amounts of unlabelled data in order to learn about the underlying predictive regularities.

Using the definitions described in Section 2, the original FRFS approach can be altered to handle both labelled and unlabelled data. Consider a feature subset $B \subseteq \mathcal{A}$, the membership degree to the positive region is computed as defined in Section 2. For the semi-supervised paradigm, some data instances in X are missing. Each of these data instances is considered to belong to its own decision class, which contains only that data instance. This impacts the calculation of the positive region (10).

Theorem 1. *Let L be a set of labelled instances and U a set of unlabelled instances, with $L \cap U = \emptyset$ and $L \cup U = X$. The membership degree of an instance $x \in X$ to the positive region in the system can be defined as:*

$$POS_B^{ssl}(x) = \begin{cases} \inf\limits_{y \neq x} (1 - R_B(x, y)) & \text{if } x \in U \\ \inf\limits_{y \in (U \cup co([x]_d^L))} (1 - R_B(x, y)) & \text{if } x \in L, \end{cases}$$

where $[x]_d^L$ denotes the set of labelled instances with the same decision value as x and $co(\cdot)$ is the complement operator.

Proof. First consider $x \in U$. Its decision class contains only x, implying that (10) immediately simplifies to

$$POS_B^{ssl}(x) = \inf\limits_{y \neq x} (1 - R_B(x, y)).$$

Now assume $x \in L$. The decision class of x consists of all labelled instances y with $d(x) = d(y)$. All unlabelled instances are not part of it, as they each belong to their own, distinct classes. In (10), the infimum is therefore taken over $U \cup co([x]_d^L)$, i.e.,

$$POS_B^{ssl}(x) = \inf\limits_{y \in (U \cup co([x]_d^L))} (1 - R_B(x, y)).$$

$\square$

When $L \neq \emptyset$ and unlabelled data is considered, the dependency degree is defined as

$$\gamma_B^{ssl} = \frac{\sum\limits_{x \in X} POS_B^{ssl}(x)}{|X|}.$$

Theorem 2. *For every $B \subseteq \mathcal{A}$, $\gamma_B^{ssl} \leq \gamma_B$.*

Proof. In general, given a function f and sets S and T with $S \subseteq T$,

$$\inf\limits_{x \in T} f(x) \leq \inf\limits_{x \in S} f(x)$$

always holds, as the infimum on the left-hand side is taken over a larger set.

Consider $x \in X$. In the semi-supervised paradigm, either $x \in U$ or $x \in L$ holds. In the first case, it can be found that:

$$
\begin{aligned}
POS_B^{ssl}(x) &= \inf_{y \neq x}(1 - R_B(x, y)) \\
&= \inf_{y \in (X \setminus \{x\})}(1 - R_B(x, y)) \\
&\leq \inf_{y \in co([x]_d)}(1 - R_B(x, y)) \\
&= \inf_{y \notin [x]_d}(1 - R_B(x, y)) = POS_B(x)
\end{aligned}
\tag{12}
$$

where $[x]_d$ is determined within the fully labelled system. Step (12) holds since $co([x]_d) \subseteq (X \setminus \{x\})$. Secondly, if $x \in L$, then

$$
\begin{aligned}
POS_B^{ssl}(x) &= \inf_{y \in (U \cup co([x]_d^L))}(1 - R_B(x, y)) \\
&\leq \inf_{y \in co([x]_d)}(1 - R_B(x, y)) \\
&= \inf_{y \notin [x]_d}(1 - R_B(x, y)) = POS_B(x)
\end{aligned}
\tag{13}
$$

where $[x]_d$ is determined within the fully labelled data. Step (13) holds since $co([x]_d) \subseteq (U \cup co([x]_d^L))$. Indeed, every data instance $y \in co([x]_d)$ remains either labelled in the semi-supervised paradigm $(y \in co([x]_d^L))$ or has no label $(y \in U)$. It can therefore be concluded that

$$
(\forall x \in X)(POS_B^{ssl}(x) \leq POS_B(x)).
\tag{14}
$$

Finally, it can be found that:

$$
\gamma_B^{ssl} = \frac{\sum_{x \in X} POS_B^{ssl}(x)}{|X|} \overset{(14)}{\leq} \frac{\sum_{x \in X} POS_B(x)}{|X|} = \gamma_B.
$$

$\square$

This dependency measure can be used in the same way as the original definition as a basis for guiding the search toward optimal subsets. The modified greedy hillclimbing search method ssFRFS is presented in Algorithm 1. The algorithm iteratively adds the best feature (based on γ^{ssl}) to the current subset until the maximal value is attained for the dataset. At this point, the subset is a (super)reduct, because Theorem 2 shows that maximizing γ^{ssl} implies maximizing γ as well.

4 Experimental Evaluation

This section details the experiments conducted and the results obtained for the novel ssFRFS approach. The proposed method was applied to 12 benchmark

---
Algorithm 1: The ssFRFS algorithm
---
input : $\mathcal{A}$, the set of all conditional features

output: B, the reduct

1 $B \leftarrow \{\}; \gamma_{best}^{ssl} = 0$

2 **while** $\gamma_{best}^{ssl} \neq \gamma_{\mathcal{A}}^{ssl}$ **do**

3 $\quad T \leftarrow B$

4 $\quad$ **foreach** $a \in (\mathcal{A} \setminus B)$ **do**

5 $\quad\quad$ **if** $\gamma_{B \cup \{a\}}^{ssl} > \gamma_T^{ssl}$ **then**

6 $\quad\quad\quad T \leftarrow B \cup \{a\}$

7 $\quad\quad\quad \gamma_{best}^{ssl} = \gamma_T^{ssl}$

8 $\quad B \leftarrow T$

9 **return** B
---

datasets of different sizes. The results presented here relate to the performance in terms of quality of subsets obtained: classification accuracy and subset size. The effect of the random removal of 10-90% of the class labels from the datasets on the proposed approach is also assessed.

4.1 Experimental Setup

The 12 different benchmark datasets are drawn from [4]. The class labels are randomly removed from 10%, 30%, 50%, 70% and 90% of the labelled data for each dataset in order to simulate the semi-supervised problem domain. ss-FRFS is compared with the original fuzzy-rough feature selection using a greedy hill-climbing search [5] and applied to the original, fully-labelled data. For all approaches, the Łukasiewicz t-norm $(\max(x + y - 1, 0))$ and the Łukasiewicz fuzzy implicator $(\min(1 - x + y, 1))$ are adopted to implement the fuzzy connectives in (7) and (8).

For the generation of classification results, two different classifier learners have been employed: the rule-based classifier JRip [1] and a nearest-neighbour classifier IBk (with $k = 3$). Five stratified randomisations of 10-fold cross-validation were employed in generating the classification results. It is important to point out here that FS is performed as part of the cross-validation and each fold results in a new selection of features. For the comparison of ssFRFS in terms of model performance, average classification accuracy is used. In addition, a non-parametric Wilcoxon test is performed [10] (significance level $\alpha = 0.05$) in order to evaluate whether any differences in performance are statistically significant. In particular, FRFS is compared with ssFRFS, in order to verify whether the presence of unlabelled data implies a reduction in performance.

4.2 Results

The results of the experimental evaluation are shown in Tables 1-3. The results in Table 1 show that the approach achieves a reduction in the subset sizes for

almost all cases, with a few small notable exceptions of the *australian* and *glass* datasets when the number of missing labels is greater than 70%. The *segment* dataset also shows a similar pattern starting at 30%. It should be noted however that the corresponding level of reduction using FRFS with no missing labels is very small - only a single feature, so this is expected to some degree.

Tables 2 and 3 detail the classification results for the JRip and IBk classifier learners respectively. Examining these results, it is clear that even for those cases with high levels of label removal, ssFRFS can still return comparable classification accuracies to the case where no labels have been removed. Clearly, it would be unrealistic to expect that this approach would be able to maintain the robustness of the fully supervised data (no missing labels). After all, there is much discriminative information encoded in class labels, but overall ssFRFS performs well. There is a small decrease in performance when compared to the unreduced data, but the statistical analysis in Tables 4 and 5 shows that no significant differences are observed when the results are compared with the fully supervised approach, regardless of the level of missing labels.

Dataset	No. of features	No. of objects	FRFS	Missing class labels				
				10%	30%	50%	70%	90%
australian	14	690	12.90	13.12	13.70	13.96	14.00*	14.00*
cleveland	13	303	7.64	8.32	9.20	9.78	10.08	10.24
ecoli	7	336	6.00	6.00	6.00	6.00	6.00	6.00
glass	9	214	9.00	8.84	8.92	8.98	9.00*	9.00*
heart	13	270	7.06	7.96	8.90	9.72	10.14	10.32
ionosphere	34	351	7.04	12.42	16.18	17.32	18.48	18.60
olitos	25	120	5.00	5.10	5.44	5.74	5.86	5.94
segment	18	2310	16.00	17.98	18.00*	18.00*	18.00*	18.00*
vehicle	18	946	8.42	9.74	11.04	11.46	11.70	11.84
water2	38	390	6.00	6.46	6.92	7.00	7.04	7.10
water3	38	390	6.00	6.26	6.94	7.00	7.02	7.00
wine	13	178	8.42	5.80	6.34	6.56	6.74	6.80

* denotes that no reduction took place

Table 1: Average subset size in terms of original number of features.

5   Conclusion

This paper has presented a new approach to feature selection for semi-supervised data. One of the primary motivations behind the development of the approach was to propose a means of carrying out FS when some of the data class labels are missing. Indeed, there is no bound on the number of labelled data objects required and even if all labels are missing, the approach will still return a valid fuzzy-rough (super)reduct. This appealing property means that no consideration

Dataset			Missing class labels				
	Unred.	FRFS	10%	30%	50%	70%	90%
australian	85.16	85.16	84.87	85.13	85.10	85.1	85.16
cleveland	54.23	54.48	54.34	55.10	54.63	54.69	54.42
ecoli	82.26	81.13	81.13	81.13	81.13	81.13	81.13
glass	67.17	67.17	66.41	66.99	66.54	67.17	67.17
heart	72.96	74.15	74.37	74.67	75.04	75.04	75.93
ionosphere	87.57	87.74	85.13	86.52	85.39	86.96	86.61
olitos	68.50	62.83	59.83	59.83	63.67	62.00	61.67
segment	95.31	95.11	95.31	95.28	95.28	95.28	95.28
vehicle	68.65	63.50	65.72	65.58	66.58	66.10	67.07
2-completed	82.15	83.28	82.97	83.08	82.56	81.54	82.67
3-completed	82.72	81.23	79.79	79.85	78.36	77.64	78.72
wine	93.54	91.46	89.44	87.08	87.01	86.91	87.31

Table 2: Classification accuracy using JRip

needs to be given to any tunable parameters in order to account for differing levels of supervision of the data. Indeed the experimental evaluation has demonstrated that the approach is particularly useful even with high levels of missing labels.

The ideas described in this paper offer some new directions for further development. In particular, other supervised fuzzy-rough approaches could be extended in a similar way to that proposed in this paper, including the fuzzy discernibility matrix-based approaches described in [5]. Also, the idea of object weighting could be incorporated [9], where objects with missing labels are considered to be less important in the calculations when compared to those that are labelled.

The basic semi-supervised concepts could be further generalised to the situation where both conditional and decision features may be missing from the data. This could form the basis for a number of approaches for performing feature selection on sparse data. For the work described here, the fuzzy connectives and similarities are the same as those used for the typical supervised approaches, but it would be interesting to investigate the influence of different choices in this regard.

Acknowledgment

Dataset	Unred.	FRFS	Missing class labels				
			10%	30%	50%	70%	90%
australian	81.62	79.57	79.39	80.81	81.22	81.62	81.62
cleveland	55.11	49.24	51.93	53.08	55.23	56.51	56.78
ecoli	80.78	80.49	80.49	80.49	80.49	80.49	80.49
glass	70.68	70.68	70.40	70.49	70.49	70.68	70.68
heart	77.11	72.96	75.19	75.11	76.89	78.52	78.89
ionosphere	85.30	86.43	84.52	85.22	84.43	85.57	84.52
olitos	76.33	63.67	63.33	63.33	64.17	64.00	63.17
segment	97.13	97.57	97.13	97.13	97.13	97.13	97.13
vehicle	70.00	63.24	64.82	66.22	66.05	66.64	67.06
2-completed	85.03	83.59	81.23	80.31	79.90	79.95	78.77
3-completed	80.77	78.56	76.92	76.36	76.46	73.64	74.72
wine	95.07	95.41	92.56	92.47	92.48	92.05	91.48

Table 3: Classification accuracy using IBk

FRFS+IBk (0%)	R^+	R^-	P-value
ssFRFS+IBk (10%)	43.0	23.0	≥ 0.2
ssFRFS+IBk (30%)	35.0	31.0	≥ 0.2
ssFRFS+IBk (50%)	31.0	35.0	≥ 0.2
ssFRFS+IBk (70%)	36.5	41.5	≥ 0.2
ssFRFS+IBk (90%)	40.5	37.5	≥ 0.2

Table 4: Comparison of FRFS with ssFRFS for IBk

References

1. W.W. Cohen. Fast effective rule induction, Proceedings of the 12th International Conference on Machine Learning, pp. 115–123, 1995.
2. C. Cornelis, R. Jensen, G. Hurtado Martín, D. Ślęzak. Attribute Selection with Fuzzy Decision Reducts, Information Sciences, vol. 180, no. 2, pp. 209–224, 2010.
3. D. Dubois and H. Prade. Putting Rough Sets and Fuzzy Sets Together, *in* Intelligent Decision Support, pp. 203–232, 1992.
4. A. Frank and A. Asuncion. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2010.
5. R. Jensen and Q. Shen. New Approaches to Fuzzy-Rough Feature Selection, IEEE Transactions on Fuzzy Systems, vol. 17, no. 4, pp. 824–838, 2009.
6. R. Jensen, A. Tuson, and Q. Shen. Finding rough and fuzzy-rough set reducts with SAT, Information Sciences, vol. 255, pp. 100–120, 2014.
7. Z. Pawlak. Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishing, 1991.
8. A.M. Radzikowska and E.E. Kerre. A comparative study of fuzzy rough sets, Fuzzy Sets and Systems, vol. 126, no. 2, pp. 137–155, 2002.
9. S. Widz and D. Ślęzak. Attribute subset quality functions over a universe of weighted objects. Proceedings of RSEISP 2014, pp. 99-110, 2014.

FRFS+JRip (0%)	R^+	R^-	P-value
ssFRFS+JRip (10%)	52.0	14.0	0.10156
ssFRFS+JRip (30%)	44.0	22.0	≥ 0.2
ssFRFS+JRip (50%)	38.0	28.0	≥ 0.2
ssFRFS+JRip (70%)	41.5	24.5	≥ 0.2
ssFRFS+JRip (90%)	42.5	23.5	≥ 0.2

Table 5: Comparison of FRFS with ssFRFS for JRip

10. F. Wilcoxon. Individual comparisons by ranking methods, Biometrics bulletin, vol. 1, no. 6, pp. 80–83, 1945.