



HAL
open science

ROSES: A Novel Semi-supervised Feature Selector

Xiaoyu Zhang, Keyu Liu, Jing Ba, Xin Yang, Xibei Yang

► **To cite this version:**

Xiaoyu Zhang, Keyu Liu, Jing Ba, Xin Yang, Xibei Yang. ROSES: A Novel Semi-supervised Feature Selector. 12th International Conference on Intelligent Information Processing (IIP), May 2022, Qingdao, China. pp.336-347, 10.1007/978-3-031-03948-5_27 . hal-04178734

HAL Id: hal-04178734

<https://inria.hal.science/hal-04178734v1>

Submitted on 8 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

ROSES: A novel semi-supervised feature selector

Xiaoyu Zhang¹, Keyu Liu^{1,2}, Jing Ba¹, Xin Yang³, and Xibei Yang^{1*}

¹ School of Computer, Jiangsu University of Science and Technology, Zhenjiang
212003, China

² School of Computing and Artificial Intelligence, Southwest Jiaotong University
Chengdu 611756, China

³ School of Economic Information Engineering, Southwestern University of Finance
and Economics, Chengdu 611130, China

Abstract. In this study, a novel ROugh set based Semi-supervised fEature Selector (ROSES) was developed to pre-process partially labeled data. The main innovations of our ROSES are: 1) the selected features over labeled samples laid the foundation for further searching qualified features over unlabeled samples; 2) a new granularity related measure was designed to quickly evaluate features. Through testing four different ratios (20%, 40%, 60%, 80%) of labeled samples, the experimental results over 15 UCI datasets demonstrated that our framework is superior to the other five popular partially labeled data feature selectors: 1) the feature subsets identified by ROSES offer competitive classification performances; 2) ROSES is good at seeking a balance between efficiency of searching features and effectiveness of the selected features.

Keywords: Feature selection · Granularity · Partially labeled · ROSES · Rough sets.

1 Introduction

Attribute reduction is the key in the field of rough set [1]. It can be regarded as a valuable technique to perform feature selection. Though such a technique has been conventionally explored over samples with complete labels, i.e., labeled data, much attention has recently been paid to samples unlabeled data or samples with missing labels, which are required for practical applications.

Generally speaking, a data contains samples with missing labels can be termed as the partially labeled data [2, 3]. Such a type of data is more complicated than labeled data and unlabeled data. As far as the topic of feature selection is concerned, it may be argued that both the techniques related to labeled and unlabeled data should be fully taken into account if partially labeled data is studied.

Presently, many pioneering researchers have developed various rough set based feature selection strategies over partially labeled data. For example, Dai et al. [4] have proposed a semirough-P which fuses the evaluations of features over

* Corresponding author: jsjxy_yxb@just.edu.cn

both labeled and unlabeled samples; Pang and Zhang [5] have introduced the neighborhood discrimination index [6] into partially labeled data, which aims to evaluate the significance of features more exactly; Liu et al. [7] have completed the missing labels in partially labeled data by using LPA [8], and then introduced an ensemble selector [9] into such a revised data for evaluate features.

Nevertheless, most of the previous studies about feature selection over partially labeled data are focused on evaluating features by either new measures or new angles. In other words, the keys of the employed feature selectors are similar to those which have been exploited over labeled and unlabeled data. From this point of view, the difference between the structures inherent in labeled and unlabeled samples may be overlooked. For instance, in Ref. [7], LPA is used to complete the labels of unlabeled samples, it follows that the raw structure of unlabeled samples may be disturbed because the performance of LPA relies heavily on the structure of labeled data.

From discussions above, we want to work out how to make full use of labeled and unlabeled data while maintaining a certain speed, in the context of this study, a novel feature selection framework called ROugh set based Semi-supervised fEature Selector (ROSES) will be developed. Different from existing approaches, our framework divides the whole process of feature selection over partially labeled data into two phases: 1) selecting qualified features over labeled samples and then derive a temporary result; 2) based on the information provided by unlabeled samples, further adding qualified features into such a temporary result.

Obviously, the above two phases imply an efficient guidance mechanism which is useful in speeding up the process of feature selection over partially labeled data. Furthermore, a new measure related to the concept called granularity is proposed in the second stage of searching. The calculation of such a measure is low cost and it can also be used to quickly identify qualified features over unlabeled samples. Such two points demonstrate the main contributions of our ROSES.

The remainder of this paper is organized as follows. In Section 2, we will briefly review some basic notions related to partially labeled data. Our framework are introduced in Section 3. In Section 4, comparative experimental results are presented, as well as corresponding analyses. The paper is ended with conclusions and perspectives for future work in Section 5.

2 Preliminary knowledge

In this section, we will briefly review some basic notions related to partially labeled data. Without loss of generality, a partially labeled data can be denoted by $DS = \langle U^l \cup U^{ul}, AT, d^l \rangle$: U^l is a set of labeled samples; U^{ul} is a set of unlabeled samples; AT is a set of features for characterizing samples in both U^l and U^{ul} ; d^l is used to record the labels of samples in U^l .

In rough set, an indiscernibility relation or some other binary relations are frequently constructed for the convenience of granulating samples [10], which are

based on the information provided by features. For instance, Hu et al. [11] have developed a neighborhood relation: $\delta_{AT}^U = \{(x_i, x_j) \in U \times U : dis_{AT}(x_i, x_j) \leq \delta\}$ in which $dis_{AT}(x_i, x_j)$ is the distance between samples x_i and x_j over AT , $U \subseteq U^l \cup U^{ul}$, $\delta \geq 0$ is a given radius.

Following $\delta_{AT}^U, \forall x_i \in U$, the neighborhood of x_i is then obtained by $\delta_{AT}^U(x_i) = \{x_j \in U : (x_i, x_j) \in \delta_{AT}^U\}$. Such a process can be regarded as an information granulation over U . To further characterize the degree of information granulation, the following definition of granularity can be used.

Definition 1. *Given a partially labeled data DS , $\delta \geq 0$ is a radius, the neighborhood relation related granularity over $U \subseteq U^l \cup U^{ul}$ is:*

$$G_{AT}^\delta(U) = \frac{|\delta_{AT}^U|}{|U|^2} = \frac{\sum_{x_i \in U} \delta_{AT}^U(x_i)}{|U|^2}, \quad (1)$$

$|X|$ denotes the cardinal number of set X .

Based on Def. 1, it is not difficult to conclude that a finer neighborhood relation indicates a small value of granularity, a coarser neighborhood relation will generate a great value of granularity.

Furthermore, by considering the labels of samples in U^l , from the viewpoint of rough set, we can induce a partition over U^l such that $U^l/IND_{d^l} = \{X_1, \dots, X_t\}$. $\forall X_s \in U^l/IND_{d^l} (1 \leq s \leq t)$, X_s is regarded as the s -th decision class which contains all samples with same label. Immediately, the following neighborhood rough set can be defined over U^l .

Definition 2. *Given a partially labeled data DS , $\forall X_s \in U^l/IND_{d^l}$, $\delta \geq 0$ is a radius, the neighborhood based lower and upper approximations of X_s are:*

$$\underline{X}_{sAT}^\delta = \{x_i \in U : \delta_{AT}^{U^l}(x_i) \subseteq X_s\}; \quad (2)$$

$$\overline{X}_{sAT}^\delta = \{x_i \in U : \delta_{AT}^{U^l}(x_i) \cap X_s \neq \emptyset\}. \quad (3)$$

Following Def. 2, a crucial concept called approximation quality [12] can be formed for the measurement of degree of certain belongingness.

Definition 3. *Given a partially labeled data DS , $\delta \geq 0$ is a radius, the neighborhood based approximation quality over U^l is:*

$$\gamma_{AT}^\delta(U^l) = \frac{|d_{AT}^{U^l}|}{|U^l|} = \frac{|\bigcup_{s=1}^t \underline{X}_{sAT}^\delta|}{|U^l|}. \quad (4)$$

Obviously, the above rough set and approximation quality are only defined over labeled samples. Furthermore, to analyze unlabeled samples in partially labeled data by rough set, the following unsupervised relevance [13] can be used.

Definition 4. *Given a partially labeled data DS , $\delta \geq 0$ is a radius, the neighborhood based unsupervised relevance over U^{ul} is defined as:*

$$\gamma_{AT}^\delta(U^{ul}) = \frac{1}{|AT|} \sum_{a \in AT} \frac{|d_a^{ul\delta}|}{|U^{ul}|}, \quad (5)$$

d_a^{ul} is used to record the pseudo labels of unlabeled samples in U^{ul} , which is generated by single attribute a .

In Def. 4, each single attribute is required to generate the pseudo labels of unlabeled samples in U^{ul} and then the neighborhood relation $\delta_{AT}^{U^{ul}}$ can be constructed for rough characterization of decision classes related to those pseudo labels. Therefore, the number of approximation qualities which will be generated is $|AT|$, a mean value of those approximation qualities is regarded as the unsupervised relevance.

3 Proposed framework

Following the above section, it is not difficult to draw the following conclusions.

- 1) Approximation quality shown in Def. 3 can be directly used to select qualified features over labeled samples. In such a mechanism, the value of approximation quality related to AT is regarded as the feature evaluation criterion and then features which provide higher value of approximation quality will be selected.
- 2) Unsupervised relevance shown in Def. 4 can be employed to construct feature selection criterion for selecting qualified features over unlabeled samples. In such a mechanism, features which provide higher value of unsupervised relevance will be selected.

However, our objective in this study is to select qualified features over partially labeled samples instead of mere labeled samples or unlabeled samples, and then we have been faced with the problem of making a bridge between the above two mechanisms of selecting features. To this end, we will design a novel framework to perform feature selection over partially labeled data. The detailed steps of our framework are shown as follows.

- 1) Select qualified features over labeled samples based on measure shown in Def. 3. In this case, the widely accepted forward greedy searching [14] can be employed.
- 2) Determine whether the above selected features satisfies the criterion related to unsupervised relevance shown in Def. 4. If so, then terminate the procedure; otherwise, go to next step.
- 3) By calculating unsupervised relevances in terms of all candidate features, select some other valuable features step by step until the required criterion is satisfied.

From discussions above, a detailed algorithm can be designed as follows. The main structures of our algorithm is categorized into the following two aspects: 1) labeled and unlabeled samples in partially labeled data are processed separately; 2) the selected features over labeled samples may guide a further searching of features over unlabeled samples.

Furthermore, in the process of searching features over unlabeled samples, a new measure $\frac{G_a^\delta(U^{ul})}{dis(a, c)}$ is proposed for evaluating candidate features and then identifying an appropriate feature for each iteration. By comparing with unsupervised relevance, the computation of such a measure is simple and low cost. This is mainly because the time complexity of computing $\frac{G_a^\delta(U^{ul})}{dis(a, c)}$ is $\mathcal{O}(|U^{ul}|^2)$, the time complexity of computing unsupervised relevance is $\mathcal{O}(|U^{ul}|^2 \cdot |A|^2)$ in which $|A|$ keeps increasing with the increasing of the number of iterations.

Algorithm 1: ROugh set based Semi-supervised fEature Selection (ROSES).

Input: A Decision system DS , radius $\delta \geq 0$.
Output: One subset A of selected features.

- 1 $A = \emptyset$;
- // **Select qualified features over labeled samples**
- 2 Calculate $\gamma_{AT}^\delta(U^l)$;
- 3 $\gamma_A^\delta(U^l) = -\infty$;
- 4 **while** $\gamma_A^\delta(U^l) < \gamma_{AT}^\delta(U^l)$ **do**
- 5 $\forall a \in AT - A$, evaluate a by calculating $\gamma_{A \cup \{a\}}^\delta(U^l)$;
- 6 Select a feature $b \in AT - A$ with the maximal value of evaluation;
- 7 $A = A \cup \{b\}$;
- 8 Calculate $\gamma_A^\delta(U^l)$;
- 9 **end**
- // **Select qualified features over unlabeled samples**
- 10 Calculate $\gamma_A^\delta(U^{ul})$ and $\gamma_{AT}^\delta(U^{ul})$;
- 11 **if** $\gamma_A^\delta(U^{ul}) \geq \gamma_{AT}^\delta(U^{ul})$ **then**
- 12 go to step 21;
- 13 **else**
- 14 **while** $\gamma_A^\delta(U^{ul}) < \gamma_{AT}^\delta(U^{ul})$ **do**
- 15 $\forall a \in AT - A$, evaluate a by calculating $\frac{G_a^\delta(U^{ul})}{dis(a, c)}$;
- // c is the latest feature in A .
- // $dis(a, c)$ is the distance between two features.
- 16 Select a feature $b \in AT - A$ with the maximal value of evaluation;
- 17 $A = A \cup \{b\}$;
- 18 Calculate $\gamma_A^\delta(U^{ul})$;
- 19 **end**
- 20 **end**
- 21 **return** A ;

4 Experiments

In this section, all experiments are conducted on a personal computer powered with Window 10, intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz and 16.00GB memory. Additionally, 15 public data sets from UCI Machine Learning Repository have been employed to conduct the experiments. The detailed statistics of these data are shown in the following Tab. 1.

Table 1: Details of data sets

ID	Data sets	# Samples	# Features	# Labels
1	Breast Cancer Wisconsin (Diagnostic)	569	30	2
2	Cardiotocography	2126	21	10
3	Diabetic Retinopathy Debrecen	1151	19	2
4	Forest Type Mapping	523	27	4
5	German	1000	24	2
6	Ionosphere	351	34	2
7	Libras Movement	360	90	15
8	Musk (Version 1)	476	166	2
9	Parkinson Multiple Sound Recording	1208	26	2
10	QSAR Biodegradation	1055	41	2
11	Seeds	210	7	3
12	Statlog (Image Segmentation)	2310	18	7
13	Statlog (Vehicle Silhouettes)	846	18	4
14	Ultrasonic Flowmeter Diagnostics-Meter D	180	43	4
15	Urban Land Cover	675	147	9

To ensure the reliability and stability of comparative studies, 10-fold cross validation is used. That is, given one data set, it is divided into 10 groups with same size. Then, for each round of computation, 9 groups compose the training data and the remainder is taken as the testing data. Furthermore, for each training data, we have also set the proportions of unlabeled samples as 20%, 40%, 60% and 80%, respectively.

Regarding the parameter of neighborhood rough set, we have appointed 20 different radii such that $0.02, 0.04, \dots, 0.4$ for calculating granularity, approximation quality and unsupervised relevance.

In what follows, our algorithm (ROSES) will be compared with other five state-of-the-art methods, they are: 1) semirough-P [4] (Attribute Selection for Partially Labeled Data); 2) RSES [7] (Rough set based Semi-supervised feature selection via Ensemble Selector); 3) FW-SemiFS [15] (Forward Semi-supervised Feature Selection); 4) RRPC [16] (Max-Relevance and Min-Redundancy via Pearsons Correlation Coefficient); 5) SSNDI [5] (Semi-supervised Neighborhood Discrimination Index).

4.1 Comparisons among classification accuracies

In this subsection, classification accuracies related to the selected features will be compared among six different algorithms. To obtain the classification results over partially labeled data, a graph-based semi-supervised classifier [17] was used. When used, the test data is imported into the semi-supervised classifier along with the results of the training data. The detailed results are shown in the following Figs. 1-4.

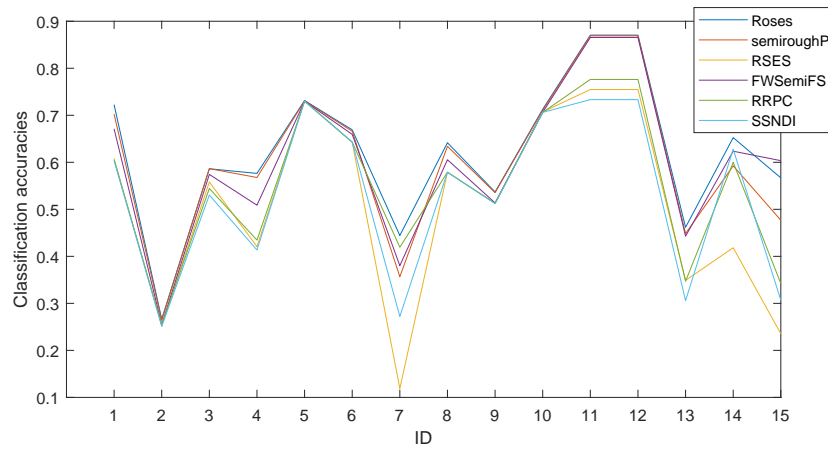


Fig. 1: Classification accuracies (20% of unlabeled samples)

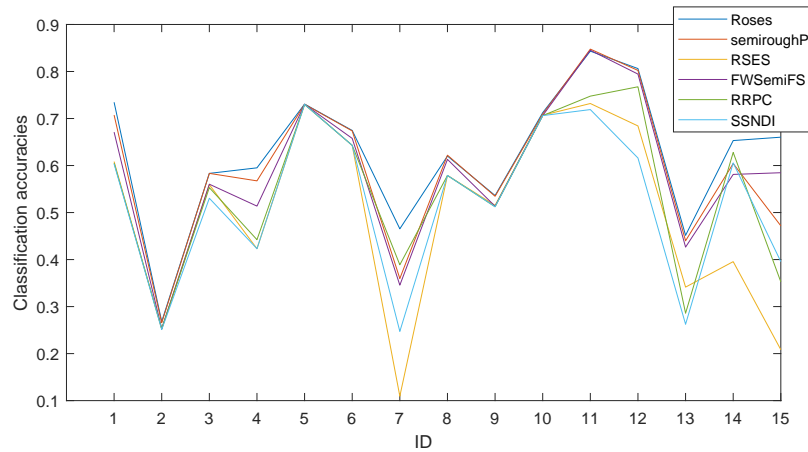


Fig. 2: Classification accuracies (40% of unlabeled samples)

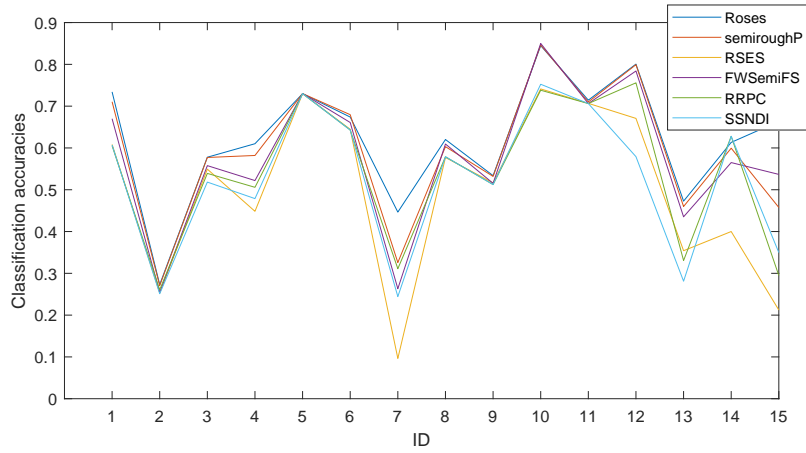


Fig. 3: Classification accuracies (60% of unlabeled samples)

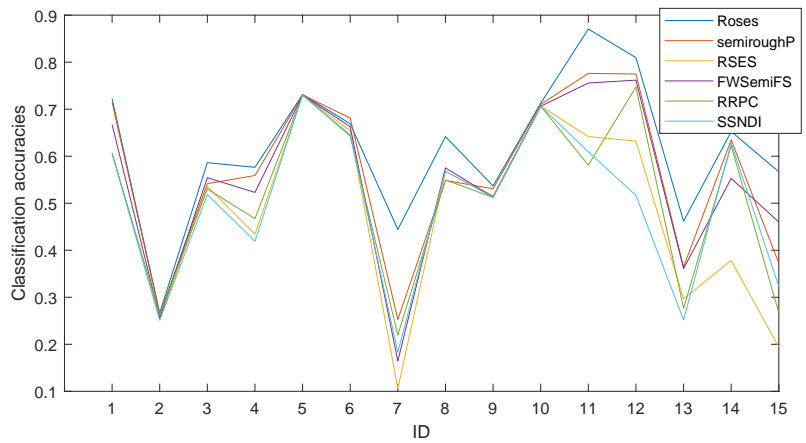


Fig. 4: Classification accuracies (80% of unlabeled samples)

With a thorough investigation, it is not difficult to reveal: for the four different ratios of unlabeled samples we have tested, because we made full use of the labeled data, and then processed the unlabeled data based on it features identified by our approach provide competitive classification performances and have good classification ability for different data.

4.2 Comparisons among time consumptions

Besides the effectiveness of selected features, efficiency of the process of selecting features is another index which should be paid much attention to. Therefore, the following Tabs. 5-8 show detailed time consumptions of searching features in terms of 6 different algorithms.

Table 2: Elapsed time (Seconds, 20% of unlabeled samples)

ID	ROSES	semirough-P	RSES	FW-SemiFS	RRPC	SSNDI
1	9.8410	34.7805	2.3042	807.9025	0.0473	7.8902
2	1.8457	6.4650	0.4296	140.2531	0.0131	0.7719
3	1.5791	5.5891	0.1473	57.3781	0.1156	0.6514
4	2.3610	8.4731	0.1828	131.4534	0.0247	1.0268
5	3.9058	45.5061	0.2805	187.6341	0.2571	2.1024
6	1.0944	6.3066	0.0436	28.9118	0.0236	0.1971
7	4.9920	24.3230	0.1202	257.9602	0.0631	0.9288
8	3.6975	15.2654	0.9464	251.5692	0.0162	1.1063
9	5.7617	25.9482	1.7091	371.5844	0.0192	1.4837
10	9.0672	33.5578	3.8300	892.9948	0.0312	7.0059
11	0.0660	0.1449	0.0141	2.5931	0.0094	0.1048
12	0.9459	3.1381	0.2291	81.8578	0.0221	0.6346
13	1.0135	6.0669	0.0860	14.4555	0.1107	0.2025
14	8.3566	46.5414	0.3952	656.1304	0.2852	7.4682
15	1.2416	4.9355	0.0931	36.9320	0.0139	0.4151
Average	3.7180	17.8028	0.7208	261.3074	0.0702	2.1326

Table 3: Elapsed time (Seconds, 40% of unlabeled samples)

ID	ROSES	semirough-P	RSES	FW-SemiFS	RRPC	SSNDI
1	6.7874	21.8645	3.5453	810.5540	0.0410	12.2549
2	1.1446	3.8866	0.5136	95.7897	0.0053	0.7390
3	0.9715	3.2948	0.0985	32.2797	0.0179	0.4251
4	1.3665	5.2340	0.1967	93.0061	0.0063	0.9060
5	3.8372	45.5269	0.2033	153.5287	0.3036	2.8257
6	0.8592	4.5758	0.0368	20.7222	0.0147	0.2157
7	4.4509	38.7812	0.1478	199.8708	0.0472	0.8501
8	2.0530	8.6184	0.8815	180.8451	0.0075	1.2125
9	4.0983	16.5857	1.7082	272.5337	0.0150	1.3748
10	5.2193	17.1016	3.9603	699.0257	0.0233	8.3030
11	0.0607	0.1294	0.0106	2.3041	0.0028	0.0191
12	0.6514	2.1407	0.2227	63.4573	0.0116	0.5791
13	0.8384	4.9821	0.0677	12.8724	0.0220	0.1150
14	7.6571	57.9956	0.4045	402.2444	0.2343	7.4535
15	0.9882	3.7313	0.0941	29.9364	0.0081	0.4103
Average	2.7322	15.6299	0.8061	204.5980	0.0507	2.5122

Table 4: Elapsed time (Seconds, 60% of unlabeled samples)

ID	ROSES	semirough-P	RSES	FW-SemiFS	RRPC	SSNDI
1	5.1207	12.9354	3.7267	601.2517	0.0367	11.5602
2	0.8078	2.4154	0.4824	60.4513	0.0051	0.6070
3	0.9002	2.7397	0.0981	28.3272	0.0143	0.3110
4	1.0775	3.8580	0.5605	67.3118	0.0063	0.9177
5	4.7934	35.7112	0.1810	125.8219	0.2391	1.4298
6	0.9625	4.4285	0.0398	21.3258	0.0115	0.1472
7	4.6988	50.0832	0.1902	169.6717	0.0444	0.7432
8	1.4130	5.7734	0.9541	113.1499	0.0076	1.2112
9	3.6194	13.3789	2.1563	198.5682	0.0127	1.7350
10	3.9326	8.9997	3.5265	478.6376	0.0200	6.0474
11	0.0643	0.1185	0.0107	2.1276	0.0030	0.0182
12	0.5153	1.4163	0.2095	46.0624	0.0113	0.6397
13	0.7623	4.3229	0.0526	11.3633	0.0236	0.1150
14	7.9060	47.9627	0.3922	315.8386	0.2555	7.1038
15	0.9295	3.1501	0.0954	24.9964	0.0085	0.4483
Average	2.5002	13.1529	0.8451	150.9937	0.0467	2.2023

Table 5: Elapsed time (Seconds, 80% of unlabeled samples)

ID	ROSES	semirough-P	RSES	FW-SemiFS	RRPC	SSNDI
1	6.5311	8.9779	3.6702	386.8968	0.0336	9.5654
2	0.8539	1.6771	0.4879	29.7393	0.0052	0.6067
3	0.9414	2.4550	0.0926	20.7566	0.0149	0.2893
4	1.2292	3.4281	0.3964	50.5998	0.0059	0.8620
5	5.5429	29.2371	0.2123	82.8208	0.2144	1.3173
6	0.9992	3.7063	0.0350	16.6370	0.0086	0.1424
7	5.6846	55.8450	0.2565	140.8585	0.0444	0.7144
8	1.5969	5.1384	0.8884	70.9006	0.0074	1.1608
9	4.5283	13.6170	2.1100	149.4086	0.0136	1.6328
10	5.5023	6.4261	4.0925	324.9634	0.0230	7.1872
11	0.0764	0.1210	0.0110	1.9730	0.0031	0.0174
12	0.5845	1.3412	0.2069	21.5483	0.0112	0.5788
13	0.7890	3.8177	0.0662	9.4349	0.0232	0.1135
14	9.5079	37.7121	0.3505	228.4705	0.2183	5.8929
15	1.1010	2.9981	0.0937	19.8349	0.0075	0.3938
Average	3.0312	11.7665	0.8647	103.6562	0.0423	2.0316

Based on the above results, we can draw the following conclusions.

- 1) For the four approaches which have been constructed based on neighborhood, i.e., ROSES, semirough-P, RSES and FW-SemiFS, our ROSES ranks the second. In other words, our approach is slower than RSES while it is faster than both semirough-P and FW-SemiFS. The reason may be contributed to the fact that RSES is realized by using a measure called neighborhood decision

error rate [18]. Such a measure is over-relaxation and then fewer iterations are required to terminate RSES.

However, though RESE is efficient, the selected features by such an approach are powerless if the classification performances are concerned (see Figs. 1-4). Obviously, the corresponding classification accuracies rank behind the other comparative algorithms.

- 2) For the other two comparative methods, i.e., RRPC and SSNDI, both of them are faster than our ROSES. Firstly, RRPC uses only the Pearson's Correlation Coefficient to evaluate candidate features and then identify the qualified features, such a calculation implies a lower computational cost. Secondly, SSNDI is designed based on the measure called neighborhood discrimination index, such a measure has been demonstrate to be an efficient measure in the field of rough set [6].

Nevertheless, following Figs. 1-4, we can also observe that the effectiveness of selected features based on RRPC and SSNDI are also significantly inferior to those based on our ROSES.

5 Conclusions and future plans

In this research, we have developed a novel strategy—ROSES for searching qualified features over partially labeled data. Different from previous researches, our strategy divides the procedure of searching features into two stages. Firstly, the qualified features are selected based on conventional measures related to rough set over labeled samples. Secondly, such a subset of qualified features is further checked over unlabeled samples. If the corresponding constraint in terms of unsupervised relevance can not be satisfied, then more features will be selected by using a new granularity related measure which is equipped with lower computational cost. The framework proposed in this paper gives a new and better perspective to solve related problems in the future.

The following topics will be our further investigations.

- 1) How the ratio of unlabeled samples in data will affect the performances of our framework.
- 2) The robustness of the selected features is another interesting issue which will be addressed.

Acknowledgments

This work is supported by the Natural Science Foundation of China (Nos. 62076111, 62076088, 62006128, 61876157).

References

1. Z. Pawlak. Rough sets, *International Journal of Computer and Information Sciences* 11(3) (1982) 341-356.

2. G.B. Zhang, Z.Y. Yang, B. Huo, S.D. Chai, S. Jiang. Multiorgan segmentation from partially labeled datasets with conditional nnU-Net, *Computers in Biology and Medicine* 136 (2021) Article 104658.
3. C.Gao, J.Zhou, D.Miao, J.Wen, X.Yue. Three-way decision with co-training for partially labeled data. *Information Sciences* 544 (2021) 500-518.
4. J.H. Dai, Q.H. Hu, J.H. Zhang, H. Hu, N.G. Zheng. Attribute selection for partially labeled categorical data by rough set approach, *IEEE Transactions on Cybernetics* 47(9) (2017) 2460-2471.
5. Q.Q. Pang, Z. Li. Semi-supervised neighborhood discrimination index for feature selection, *Knowledge-Based Systems* 204 (2020) Article 106224.
6. C.Z. Wang, Q.H. Hu, X.Z. Wang, D.G. Chen, Y.H. Qian, Z. Dong. Feature selection based on neighborhood discrimination index, *IEEE Transactions on Neural Networks and Learning Systems* 29(7) (2018) 2986-2999.
7. K.Y. Liu, X.B. Yang, H.L. Yu, J.S. Mi, P.X. Wang, X.J. Chen. Rough set based semi-supervised feature selection via ensemble selector, *Knowledge-Based Systems* 165 (2019) 282-296.
8. Z.W. Zhang, X.Y. Jing, T.J. Wang. Label propagation based semi-supervised learning for software defect prediction, *Automated Software Engineering* 24 (2017) 47-69.
9. X.B. Yang, Y.Y. Yao. Ensemble selector for attribute reduction, *Applied Soft Computing* 70 (2018) 1-11.
10. Y.H. Qian, H.H. Cheng, J.T. Wang, J.Y. Liang, W. Pedrycz, C.Y. Dang. Grouping granular structures in human granulation intelligence, *Information Sciences* 382-383 (2017) 150-169.
11. Q.H. Hu, D.R. Yu, J.F. Liu, C.X. Wu. Neighborhood rough set based heterogeneous feature subset selection, *Information Sciences* 178(18) (2008) 3577-3594.
12. Z.Wu, B. Wang, N.Chen, J.Luo. Semi-monolayer covering rough set on set-valued information systems and its efficient computation, *International Journal of Approximate Reasoning* 130(2) (2021) 83-106.
13. Z. Yuan, H.M. Chen, T.R. Li, Z. Yu, B.B. Sang, C. Luo. Unsupervised attribute reduction for mixed data based on fuzzy rough sets, *Information Sciences* 572 (2021) 67-87.
14. Z. Chen, K.Y. Liu, X.B. Yang, H. Fujita. Random sampling accelerator for attribute reduction, *International Journal of Approximate Reasoning* 140 (2022) 75-91.
15. J.T. Ren, Z.Y. Qiu, W. Fan, H. Chen, P.S. Yu. Forward semi-supervised feature selection, *Proceedings of the 12th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, Osaka, Japan* (2008) 970-976.
16. X. Jin, B. Tang, H.B. He, H. Man. Semisupervised feature selection based on relevance and redundancy criteria, *IEEE Transactions on Neural Networks and Learning Systems* 28(9) (2017) 1974-1984.
17. X. Zhu, Z. Ghahramani, J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions, *Machine Learning, Proceedings of the Twentieth International Conference* (2003) 912-919.
18. Q.H. Hu, W. Pedrycz, D.R. Yu, J. Lang. Selecting discrete and continuous features based on neighborhood decision error minimization, *IEEE Transactions on Systems Man and Cybernetics Part B* 40(1) (2010) 137-150.