



A Supervised Filter Feature Selection Method for Mixed Data Based on the Spectral Gap Score

Saúl Solorio-Fernández^(✉), José Fco. Martínez-Trinidad,
and Jesús Ariel Carrasco-Ochoa

Computer Sciences Department, Instituto Nacional de Astrofísica,
Óptica y Electrónica, Luis Enrique Erro # 1, Santa María Tonantzintla,
72840 Puebla, Mexico
sausolofer@inaoep.mx

Abstract. Feature Selection for supervised classification plays a fundamental role in pattern recognition, data mining, machine learning, among other areas. However, most supervised feature selection methods have been designed for handling exclusively numerical or non-numerical data; so, in practical problems of fields such as medicine, economy, business, and social sciences, where the objects of study are usually described by both numerical and non-numerical features (mixed data), traditional supervised feature selection methods cannot be directly applied. This paper introduces a supervised filter feature selection method for mixed data based on the spectral gap score and a new kernel capable of modeling the data structure in a supervised way. To demonstrate the effectiveness of the proposed method, we conducted some experiments on public real-world mixed datasets.

Keywords: Supervised feature selection · Spectral feature selection · Mixed data · Feature ranking · Feature subset selection

1 Introduction

In areas such as pattern recognition, data mining, machine learning, statistical analysis, and in general, in tasks involving data analysis or knowledge discovery from datasets, it is common to process collections of objects¹ characterized by many features. In this situation, it is reasonable to assume that by retaining all features we will get better knowledge about the objects of study. However, in practice, this is usually not true, because many features could be either irrelevant or redundant. Indeed, it is well-known that irrelevant and redundant features may have an adverse impact on learning algorithms, decreasing the performance

¹ Also called instances, samples, tuples or observations.

of supervised classifiers and producing biases or even incorrect models [1]. Feature selection methods [2,3] have shown to be a useful tool for alleviating this problem; where the aim is to identify and eliminate irrelevant and/or redundant features in the data without significantly decreasing the prediction accuracy of a classifier built using only the selected features. Moreover, feature selection, not only reduces the dimensionality of the data facilitating their visualization and understanding; but also, it commonly leads to more compact models with better generalization ability [4].

In the literature of feature selection for supervised classification most feature selection methods (selectors) have been designed for either numeric or non-numeric data. However, these methods cannot be directly applied to mixed datasets where objects are simultaneously described by both numerical and non-numerical features. Mixed data [5] is very common, and it appears in many real-world problems; for example, in biomedical and health-care applications, socio-economic and business, software cost estimations, and so on.

In practice, for using supervised feature selection methods developed exclusively for numerical or non-numerical data on mixed data problems, it is common to apply feature transformations. The process of transforming non-numerical features to numerical ones is called encoding; nevertheless, this transformation has as a main drawback that the categories of a non-numerical feature should be coded as numerical values. This codification introduces an artificial order between feature values which does not necessarily reflect its original nature. Moreover, some mathematical operations such as addition and multiplication by a scalar do not make sense over the transformed data [6,7]. Conversely, when the feature selection methods require non-numerical data as input, an a priori data discretization for converting numerical features into non-numerical ones is needed. However, this discretization brings with it an inherent loss of information due to the binning process [7], and consequently, the results of feature selection become highly dependent on the applied discretization method. Another solution that has been considered in some supervised feature selection methods [7,8] is to analyze the numerical and non-numerical features separately, and then to merge the two set of results. However, as [5] have noted, using this solution, the associations that exist between numerical and non-numerical features are ignored.

Based on the results reported in [9], where the *spectral gap* score combined with a kernel function was successfully used for feature selection in unsupervised mixed datasets. In this paper, we propose an extension of the aforementioned method and show how using a supervised kernel and a simple leave-one-out search strategy results in a filter feature selection method useful to be applied for selecting relevant features in supervised mixed datasets. The proposed method does not transform the original space of features or process the data separately, and it can produce both a feature raking or a feature subset composed of only relevant features. Experimental results on real-world datasets show that the proposed method achieves outstanding performance compared to the state-of-the-art methods.

The rest of this paper is organized as follows. In Sect. 2, we provide a brief review of the related work, in Sect. 3, we describe the proposed method. Experiments will be presented and discussed in Sect. 4. Finally, Sect. 5 will conclude this paper and enunciate further research on this topic.

2 Related Work

In the literature, many supervised feature selection methods have been proposed [10, 11], and according to the feature selection approach they can be categorized as filter, wrapper, or hybrid. Among the more classical and relevant filter feature selection methods for supervised classification we can mention: Information Gain (IG) [12], Fisher Score [13], Gini index [14], Relief [15], and CFS [16]. IG, Fisher score, Gini index, and Relief are univariate filter methods (also called ranking-based methods) that evaluate features according to some quality criterion that quantifies the relevance of features individually; meanwhile, CFS is a multivariate feature selection method that quantifies the relevancy of features jointly, therefore it provides a feature subset as a result. IG was designed to work on non-numerical features, while Fisher score and Gini index can only process numerical features. On the other hand both CFS and Relief, according to their respective authors, can process mixed data, CFS processes numerical and non-numerical features as non-numerical (numerical features are discretized). Meanwhile, Relief deals with this problem by using the Hamming distance for non-numerical features and the Euclidean distance for numerical ones.

On the other hand, some supervised feature selection methods developed exclusively for mixed data have also been introduced, and they can be classified into four main approaches: Statistic/Probabilistic [8, 17, 18], Information Theory [7, 19–22], Fuzzy/Rough set theory [23–26], and kernel-based [27, 28] methods. In the former, the basic idea is to evaluate the relevancy of features using measures such as the joint error probability or using different correlation measures (one for each type of feature) to quantify the degree of association among features. Information theory based methods, evaluate features using measures such as mutual information or entropy. On the other hand, Fuzzy/Rough set theory methods evaluate features based on fuzzy relations or equivalence classes (also called granules). Finally, kernel-based methods perform feature selection using three components: a dedicated kernel that can handle mixed data, a feature search strategy, and a classifier (usually SVM) which allows quantifying the importance of each feature through the objective function using the dedicated kernel.

3 Proposed Method

The proposed method is inspired by the previous Unsupervised Spectral Feature Selection Method for mixed data (USFSM) introduced in [9], which is based on

Spectral Feature Selection [1]. USFSM proposes to quantify the feature consistency² by analyzing the changes in the spectrum distribution (spectral gaps) of the Symmetrical Normalized Laplacian matrix when each feature is excluded separately.

Formally, given a collection of m objects $X_F = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$, described by a set of n numerical or non-numerical features $F = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n\}$, a target concept $T = \{t_1, t_2, \dots, t_m\}$ indicating the object’s class labels, and a $m \times m$ similarity (kernel) matrix S , containing the similarities $s_{ij} \geq 0$ between all pairs of objects $\mathbf{x}_i, \mathbf{x}_j \in X_F$. Structural information from X_F can be obtained from the eigensystem of the Symmetrical Normalized Laplacian matrix $\mathcal{L}(S)$ [29] (Laplacian graph) derived from S . Specifically, the first $c + 1$ eigenvalues of $\mathcal{L}(S)$ (arranged in ascending order) and the corresponding eigenvectors contain information for separating the data X_F in c classes [1]. Nevertheless, to quantify the feature consistency using the *spectral gap* score [9] in a supervised context, we need to specify a good similarity function that uses both, the information contained the features in F , and the information provided by the target objective T . For doing this, we propose to build the object’s similarity matrix S using the following supervised kernel function:

$$s_{ij} = \frac{K(\mathbf{x}_i, \mathbf{x}_j) + D(t_i, t_j)}{2} \quad (1)$$

where $D(t_i, t_j) = 1$ if $t_i = t_j$; otherwise $D(t_i, t_j) = 0$, and $K(\mathbf{x}_i, \mathbf{x}_j)$ is the Clinical kernel defined as in [9]. With this kernel function, we are modeling the data structure in a supervised way, and at the same time, we are taking into account the contribution of each feature in the objects’ similarity.

To quantify the consistency of each feature $\mathbf{f}_i \in F$, we measure the changes that could be produced when \mathbf{f}_i is eliminated from the dataset X_F using a leave-one-out search strategy, i.e.:

$$\varphi(\mathbf{f}_i) = \gamma(X_F, c) - \gamma(X_{F_i}, c) \quad (2)$$

where X_{F_i} denotes the dataset described by the set of features F_i , which contains all features except \mathbf{f}_i , and $\gamma(\cdot, \cdot)$ is the *spectral gap* score defined as:

$$\gamma(X, c) = \sum_{i=2}^{c+1} \sum_{j=i+1}^{c+2} \left| \frac{\lambda_i - \lambda_j}{\tau} \right| \quad (3)$$

where $\lambda_i, i = 2, \dots, c+2$ are the first $c+1$ nontrivial eigenvalues of the spectrum of $\mathcal{L}(S)$, being S the similarity matrix from X , c the number of classes in the dataset, and $\tau = \sum_{i=2}^{c+2} \lambda_i$ a normalization term. In this score, the bigger the gap of the first $c+1$ eigenvalues of $\mathcal{L}(S)$, the best will be the separation between the classes.

Our proposed method begins with the construction of the similarity matrix S from the original dataset X_F along with the class label information contained

² A feature is consistent (relevant), if it takes similar values for objects that are close to each other, and takes dissimilar values for objects that are far apart.

Input : X_F : an $m \times n$ dataset, with m objects and n features
 T : label information for each object
 c : the number of classes in the dataset

Output: F_{Rank} , F_{Sub} : Feature ranking and feature subset respectively

- 1 $F = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n\}$;
- 2 Build S , the similarity matrix from X_F using (1);
- 3 Compute $\mathcal{L}(S)$ and get its spectrum;
- 4 $\gamma_F \leftarrow \gamma(X_F, c)$ // Computing the *spectral gap* score for X_F using 3;
- 5 **for** $i \leftarrow 1$ **to** n **do**
- 6 $F_i \leftarrow F \setminus \mathbf{f}_i$;
- 7 Build S' , the similarity matrix from X_{F_i} using (1);
- 8 Compute $\mathcal{L}(S')$ and get its spectrum;
- 9 Compute the feature relevancy of \mathbf{f}_i ($\varphi(\mathbf{f}_i)$) using (2);
- 10 $w[i] \leftarrow \varphi(\mathbf{f}_i)$;
- 11 **end**
- 12 Sort w in descending order and build F_{Rank} according this order;
- 13 Get the feature subset F_{Sub} consisting of those feature with $w[i] > 0$;

Algorithm 1: Supervised Spectral Feature Selection Method for mixed data (SSFMSM).

in T using 1. Then, we construct the Symmetrical Normalized Laplacian matrix $\mathcal{L}(S)$ and obtain its spectrum. Afterwards, using (3), the *spectral gap* score $\gamma(\cdot, \cdot)$ for X_F is computed. This procedure is repeated n times using a leave-one-out feature elimination strategy over F to quantify the relevance of each feature \mathbf{f}_i through (2). Finally, features in F are ranked from the most to the least consistent according to the feature weights $w[i]$, $i = 1, 2, \dots, n$ corresponding to each feature $\mathbf{f}_i \in F$, and a feature subset F_{Sub} consisting of those features with $w[i] > 0$ (relevant features) is obtained. The pseudocode of our filter method, named Supervised Spectral Feature Selection method for Mixed data (SSFMSM), is described in Algorithm 1.

4 Experiment Results

In this section, we first describe the experimental setup in Sect. 4.1, and later, the results obtained from the evaluation of the supervised filter methods over the used datasets are presented in Sect. 4.2.

4.1 Experimental Setup

To evaluate the effectiveness of our method, some experiments were done on real-world mixed datasets taken from the UCI Machine Learning repository [30]; detailed information about these datasets is summarized in Table 1. We have compared our proposal against two feature subset selectors of the state-of-the-art that can handle mixed data, namely, CFS [16] and the Fuzzy Rough Set

theory-based method introduced by Zhang et al. in [24]. Moreover, in order to contrast our method against more classical and relevant based-on-ranking filter supervised feature selection methods of the state-of-the-art, we also have made a comparison with IG [12], Relief [15], Fisher [13], and Gini index [14].

Following the standard way for assessing supervised feature selection methods, for the comparison among all selectors, we evaluate the quality of feature selection results using the classification accuracy (ACC) of the well-known and broadly used SVM [31] classifier. For the evaluation against the feature subset selectors (CFS and Zhang et al. method), we applied stratified ten-fold cross-validation, and the final classification performance is reported as the average accuracy over the ten folds. For each fold, each feature selection method is first applied on the training set to obtain a feature subset. Then, after training the classifier using the selected features, the respective test sets are used for assessing the classifier through its accuracy. Meanwhile, for the comparison against the ranking-based methods (IG, Relief, Fisher, and Gini), we compute the *aggregated accuracy* [32]. The aggregated accuracy is obtained by averaging the average accuracy achieved by the classifiers using the top $1, 2, \dots, n - 1, n$ features according to the ranking produced by each selector, in this way, we can evaluate how good is the feature ranking obtained by each selector. In order to measure the statistical significance of the results of our method, the Wilcoxon test [33] was employed over the results obtained in each dataset, and we have marked with the “+” symbol those datasets where there is a statistically significant difference of the results of the corresponding method against the results of our method.

The implementation of ranking-based methods was taken from the Matlab Feature Selection package available in the ASU Feature Selection repository [34]. Meanwhile for CFS and Zhang et al. methods we used the author’s implementation with the parameters recommended by their respective authors. For SSFSM, the Apache Commons Math³ library was used for matrix operations and eigen-system computation. All experiments were run in Matlab[®] R2018a with Java 9.04, using a computer with an Intel Core i7-2600 3.40 GHz \times 8 processor with 32 GB DDR4 RAM, running 64-bit Ubuntu 16.04 LTS (GNU/Linux 4.13.0-38 generic) operating system.

4.2 Experimental Results and Comparisons

In Tables 2 and 3, the classification accuracy reached by using the evaluated feature subset methods and the aggregated accuracy of ranking-based methods using SVM on the datasets of Table 1 are shown, respectively. In these tables, the last row shows the overall average results obtained on all tested datasets. Additionally, in the last column of Table 2, the classification accuracy using the whole set of features of each dataset is included.

As we can see in Table 2, regarding the methods for mixed data that select subsets of features, the best average results were obtained by our method, out-

³ <http://commons.apache.org/proper/commons-math/>.

Table 1. Details of the real-world mixed datasets used in our experiments.

#	Dataset	No. of objects	Features			No. of classes
			Numerical	Non-numerical	All	
1	Acute-inflammations	120	1	5	6	2
2	Automobile	205	15	10	25	6
3	Flags	194	10	19	29	8
4	Horse-colic	368	7	16	23	2
5	Post-operative	90	1	7	8	3
6	Teaching-assistant-evaluation	151	1	4	5	3
7	Thoracic-surgery	470	3	13	16	2
8	Credit-approval	690	6	9	15	2
9	Cylinder-bands	540	18	21	39	2
10	Heart-statlog	270	6	7	13	2

Table 2. Classification accuracy of SVM on the feature subsets produced by SSFSM, CFS, and Zhang et al. on the mixed datasets of Table 1.

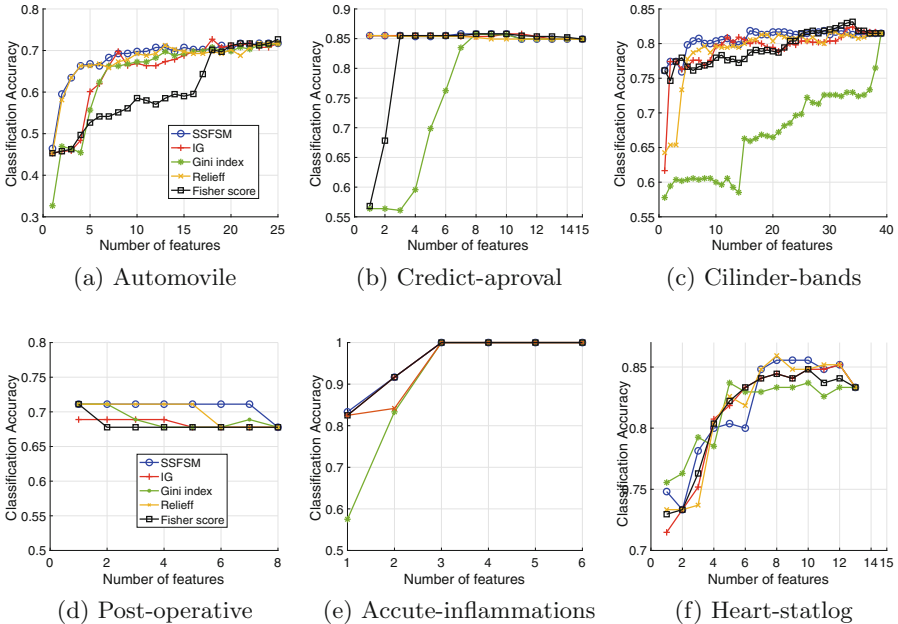
Dataset	SSFSM	CFS	Zhang et al.	Original
Acute-inflammations	1.000	1.000	1.000	1.000
Automobile	0.683	0.5892+	0.698	0.717
Flags	0.624	0.692	0.3094+	0.603
Horse-colic	0.831	0.823	0.826	0.810
Post-operative	0.711	0.678	0.689	0.678
Teaching-assistant-evaluation	0.530	0.458	0.543	0.543
Thoracic-surgery	0.845	0.845	0.845	0.845
Credit-approval	0.858	0.858	0.849	0.849
Cylinder-bands	0.804	0.7240+	0.774	0.815
Heart-statlog	0.844	0.830	0.833	0.833
Average	0.773	0.750	0.737	0.769

performing the CFS and Zhang et al. selectors, and having a statistically better performance in Automobile, Cylinder-bands, and Flags datasets. Moreover, as we can observe in this table, our method was the only one that got an average result better than those obtained using the whole set of features.

On the other hand, regarding the ranking-based methods, in Table 3, we can observe that again our method got the best results on average (average aggregated accuracy), and it was significantly better than Gini index, and Fisher score on the Acute-inflammation, Horse-colic, credit-approval, and Automovil datasets. This indicates that the ranking produced by our method is, on average,

Table 3. Aggregated accuracy of SVM on the feature ranking produced by SSFSM, IG, Gini index, Relief and Fisher score on the datasets of Table 1.

Dataset	SSFSM	IG	Gini index	Relieff	Fisher score
Acute-inflammations	0.9583	0.9569	0.9013+	0.9444	0.9569
Automobile	0.6830	0.6496	0.6419	0.6753	0.6015+
Flags	0.6223	0.6215	0.6457	0.6250	0.6320
Horse-colic	0.8253	0.8217	0.7193+	0.8266	0.8305
Post-operative	0.7069	0.6833	0.6889	0.6986	0.6819
Teaching-assistant-evaluation	0.5033	0.5062	0.4833	0.4880	0.4836
Thoracic-surgery	0.8464	0.8454	0.8507	0.8456	0.8460
Credit-approval	0.8537	0.8539	0.7598+	0.8517	0.8240
Cylinder-bands	0.8067	0.7934	0.6686	0.7905	0.7946
Heart-statlog	0.8165	0.8128	0.8145	0.8148	0.8131
Average	0.7623	0.7545	0.7174	0.7561	0.7464

**Fig. 1.** Classification accuracy of SVM on the feature ranking produced by SSFSM, IG, Gini index, Relief and Fisher score on six datasets of Table 1. The x-axis is the number of selected features. The y-axis is the classification accuracy.

better than the ranking produced by the other evaluated ranking-based selectors, as we can corroborate in most of the feature ranking plots⁴ shown in Fig. 1.

5 Conclusions and Future Work

To solve the feature selection problem in mixed data, in this paper we introduce a new supervised filter feature selection method. Our method is based on Spectral Feature Selection (*spectral gap* score) combined with a new supervised kernel and a simple leave-one-out search strategy. It results in ranking-based/feature subset selector that can effectively solve feature selection problems in supervised mixed datasets. To show the effectiveness of our method, we tested it on several public mixed datasets. The results have shown that our method, is better than CFS [16] and the method introduced by Zhang et al. in [24], two filter feature subset selectors that can handle mixed data. Furthermore, our method achieves better results than popular filter based-on raking feature selection methods of the state-of-the-art, such as IG, Relieff, Fisher score and Gini index.

A comparison using other classifiers and selectors is mandatory, and it is part of the future work of this research. Another interesting research direction is to perform a broader study about wrapper strategies that combined with our method allow to improve its performance in terms of accuracy.

Acknowledgements. The first author gratefully acknowledges to the National Council of Science and Technology of Mexico (CONACyT) for his Ph.D. fellowship, through the scholarship 224490.

References

1. Zhou, Z., Liu, H.: Spectral Feature Selection for Data Mining, pp. 1–216 (2011). 20111464
2. Liu, H., Motoda, H.: Feature Selection for Knowledge Discovery and Data Mining. Springer, Heidelberg (1998). <https://doi.org/10.1007/978-1-4615-5689-3>
3. Liu, H., Motoda, H.: Computational Methods of Feature Selection. CRC Press, Boca Raton (2007)
4. Pal, S.K., Mitra, P.: Pattern Recognition Algorithms for Data Mining, 1st edn. Chapman and Hall/CRC, Boca Raton (2004)
5. De Leon, A.R., Chough, K.C.: Analysis of Mixed Data: Methods & Applications. CRC Press, Boca Raton (2013)
6. Ruiz-Shulcloper, J.: Pattern recognition with mixed and incomplete data. Pattern Recogn. Image Anal. **18**(4), 563–576 (2008)
7. Doquire, G., Verleysen, M.: An hybrid approach to feature selection for mixed categorical and continuous data. In: Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, pp. 394–401 (2011)
8. Tang, W., Mao, K.: Feature selection algorithm for data with both nominal and continuous features. In: Ho, T.B., Cheung, D., Liu, H. (eds.) PAKDD 2005. LNCS (LNAI), vol. 3518, pp. 683–688. Springer, Heidelberg (2005). <https://doi.org/10.1007/11430919.78>

⁴ Due to limitations of space, in Fig. 1 we only show six mixed datasets of Table 1.

9. Solorio-Fernández, S., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A.: A new unsupervised spectral feature selection method for mixed data: a filter approach. *Pattern Recogn.* **72**, 314–326 (2017)
10. Kotsiantis, S.B.: Feature selection for machine learning classification problems: a recent overview. *Artif. Intell. Rev.* **42**, 157–176 (2011)
11. Tang, J., Alelyani, S., Liu, H.: Feature selection for classification: a review. In: *Data Classification*, pp. 37–64. CRC Press, 1 January 2014. <https://doi.org/10.1201/b17320>. ISBN: 9781466586741
12. Battiti, R.: Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. Neural Netw.* **5**(4), 537–550 (1994)
13. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley, Hoboken (2012)
14. Gini, C.: Variabilità e mutabilità. Reprinted in *Memorie di metodologica statistica* (Ed. Pizetti E, Salvemini, T). Libreria Eredi Virgilio Veschi, Rome (1912)
15. Kononenko, I.: Estimating attributes: analysis and extensions of RELIEF. In: Bergadano, F., De Raedt, L. (eds.) *ECML 1994*. LNCS, vol. 784, pp. 171–182. Springer, Heidelberg (1994). <https://doi.org/10.1007/3-540-57868-4-57>
16. Hall, M.A.: Correlation-based feature selection for discrete and numeric class machine learning. In: *Proceedings of the Seventeenth International Conference on Machine Learning, ICML 2000*, pp. 359–366. Morgan Kaufmann Publishers Inc., San Francisco (2000)
17. Tang, W., Mao, K.: Feature selection algorithm for mixed data with both nominal and continuous features. *Pattern Recogn. Lett.* **28**(5), 563–571 (2007)
18. Jiang, S.Y., Wang, L.X.: Efficient feature selection based on correlation measure between continuous and discrete features. *Inf. Process. Lett.* **116**(2), 203–215 (2016)
19. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(8), 1226–1238 (2005)
20. Doquire, G., Verleysen, M., et al.: Mutual information based feature selection for mixed data. In: *19th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2011)* (2011)
21. Liu, H., Wei, R., Jiang, G.: A hybrid feature selection scheme for mixed attributes data. *Comput. Appl. Math.* **32**(1), 145–161 (2013)
22. Wei, M., Chow, T.W., Chan, R.H.: Heterogeneous feature subset selection using mutual information-based feature transformation. *Neurocomputing* **168**, 706–718 (2015)
23. Hu, Q., Liu, J., Yu, D.: Mixed feature selection based on granulation and approximation. *Knowl.-Based Syst.* **21**(4), 294–304 (2008)
24. Zhang, X., Mei, C., Chen, D., Li, J.: Feature selection in mixed data: a method using a novel fuzzy rough set-based information entropy. *Pattern Recogn.* **56**, 1–15 (2016)
25. Kim, K.J., Jun, C.H.: Rough set model based feature selection for mixed-type data with feature space decomposition. *Expert Syst. Appl.* **103**, 196–205 (2018)
26. Wang, F., Liang, J.: An efficient feature selection algorithm for hybrid data. *Neurocomputing* **193**, 33–41 (2016)
27. Paul, J., Dupont, P., et al.: Kernel methods for mixed feature selection. In: *ESANN*. Citeseer (2014)
28. Paul, J., D’Ambrosio, R., Dupont, P.: Kernel methods for heterogeneous feature selection. *Neurocomputing* **169**, 187–195 (2015)
29. Luxburg, U.: A tutorial on spectral clustering. *Stat. Comput.* **17**(4), 395–416 (2007)

30. Lichman, M.: UCI Machine Learning Repository (2013)
31. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
32. Zhao, Z., Wang, L., Liu, H., Ye, J.: On similarity preserving feature selection. *IEEE Trans. Knowl. Data Eng.* **25**(3), 619–632 (2013)
33. Wilcoxon, F.: Comparisons by ranking methods. *Biometrics Bull.* **1**(6), 80–83 (1945)
34. Li, J., et al.: Feature selection: a data perspective. *ACM Comput. Surv.* **50**(6) (2017). <https://doi.org/10.1145/3136625>. ISSN: 0360-0300