I.J. Intelligent Systems and Applications, **2023, 2, 37-42**
Published Online on April 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2023.02.04

Novel Feature Selection Algorithms Based on Crowding Distance and Pearson Correlation Coefficient

Abdesslem Layeb
Constantine 2 university of Abdelhamid Mehri, NTIC faculty, LISIA laboratory
E-mail: abdesslem.layeb@univ-constantine2.dz
ORCID iD: https://orcid.org/0000-0002-6553-8253

Received: 04 April 2022; Revised: 21 August 2022; Accepted: 01 February 2023; Published: 08 April 2023

Abstract: Feature Selection is an important phase in classification models. Feature Selection is an effective task used to decrease the dimensionality and eliminate redundant and unrelated features. In this paper, three novel algorithms for feature selection problem are proposed. The first one is a filter method, the second one is a wrapper method, and the last one is a hybrid filter method. Both the proposed algorithms use the crowding distance used in the multiobjective optimization as a new metric to assess the importance of the features. The idea behind the use of the crowding distance is that the less crowded features have great impacts on the target attribute (class), and the crowded features have generally the same impact on the class attribute. To enhance the crowded distance, a combination with other metrics will give good results. In this work, the hybrid method combines between the crowding distance and Pearson correlation coefficient to well order the importance of features. Experiments on well-known benchmark datasets including large microarray datasets have shown the effectiveness and the robustness of the proposed algorithms.

Index Terms: Feature Selection, Classification, Filter Methods, Crowding Distance, Pearson Correlation.

1. Introduction

Feature selection problem is a well-known problem in data mining field. It is used in classification methods to reduce the number of features in datasets. Formally, feature selection procedure selects a subset of P significant features from a whole set of N input features, with P<N conserving a good or better accuracy compared to the entire N features [1]. It should be noted that feature selection is different from dimensionality reduction like principal component analysis (PCA). Dimensionality reduction methods create new features by combinations of attributes, whereas feature selection methods use a subset of existing attributes without changing them. Unfortunately, the feature selection problem is NP-hard problem characterized by an exponential complexity. That's why, several methods were developed to solve feature selection problem that can be regrouped in three general classes of methods: filter methods, wrapper methods and embedded methods [2].

In the filter feature selection methods, each potential feature is weighted and ranked according to a defined feature selection measure, the selection procedure consists to take the best k features. The filter methods select the potential feature independently of the classifier used. Some examples of some filter methods include the Pearson Correlation Coefficient [3], and relief feature selection [4]. The filter methods are simple, rapid, and they don't depend on any learner model. Furthermore, filter methods are easily appropriate to high-dimensional datasets. However, the greediness of the filter methods may lead to low accurate solutions, besides a redundant subset of features may be selected, in the case of highly correlated data sets, which will be unsuitable for training a classifier [2].

Unlike filter feature selection methods, Wrapper methods consider the feature selection problem as an optimization problem, where an objective function based on a predictive model is used to assess the accuracy of each selected subset of features. In this class, we can find complete search methods like branch and bound [5] or incomplete search methods like local search methods, greedy search, or metaheuristics [6, 7]. Generally, the wrapper methods give better results than the filter methods. However, there are two main drawbacks that limit the use of these methods. The first limit is the runtime complexity required for the selection. The second limit is that the performance of these methods depends on the classifier algorithm used as an objective function.

Finally, the embedded methods integrate the feature selection task in the construction process of the prediction model. In wrapper type selection methods, the classification process is divided into two parts: a learning stage and a

validation stage to validate the selected subset of features. On the other hand, the built-in methods can use all the learning examples to build the system. This is an advantage that can improve the results. Another advantage of these methods is their speed compared to wrapper approaches because they avoid the classifier to be restarted for each subset of features. Examples of embedded algorithms are the LASSO, Elastic Net and Ridge Regression [8].

As we have mentioned, filter methods are more suitable for feature selection problem. However, they depend greatly on the metric used to assess the importance of each feature. In this work, we proposed the use of a new metric for ordering the features based on the famous crowding distance used in multiobjective optimization [9,10]. The features are handled as points in a multiobjective space where the objectives are the samples. The crowding distances of the entire features are sorted in descending order. Consequently, the selected features are ranked at the top of the features ranking. The assumption behind this proposition is that the more relevant features are the most isolated feature points in the datasets space, and the more crowded features are likely to be redundant or irrelevant.

On the other hand, in the feature selection problem, Pearson correlation coefficient [11] was used to measure the importance of features and consequently remove irrelevant features. Two features are linearly correlated if their correlation coefficient is ±1; otherwise, the correlation coefficient is 0 (the features are uncorrelated). In this scope and to enhance the proposed filter crowding method, the Pearson correlation coefficient between the target feature and the remaining features are aggregated with the crowding distance leading to a more accurate filter method.

The proposed algorithms were assessed with well-known datasets and they were compared against well-known algorithms. The experimental results have proved the effectiveness of the proposed algorithms.

The rest of this paper is organized as follows. Section 2 presents the problem definition. Section 3 presents the proposed algorithms. The experiments and results are described in Section 4. Section 5 provides a conclusion.

2. Feature Selection Problem

Feature Selection problem (FS) can be defined as follows: let having a dataset with a set of N features, each feature i has a weight w_i measuring the importance of each one. The FS problem consists on choosing a subset of features that maximizes the accuracy of the model classification. The problem can be formulated as [12, 13]:

$$Maximize\ f(wixi) = \frac{Sc}{St} \tag{1}$$

$$\text{Subject :} \quad \sum_{i=1}^{N} xi \ < N \quad xi \in \{0,1\} \tag{2}$$

where "f" is a performance evaluation measure. "Sc" is the number of samples correctly classified, "St" is the total number of samples in the dataset, xi is a binary decision variable, and N is the total number of features. The objective function in this problem is to maximize the accuracy of the classification subject to one constraint, which is: the number of selected features is less than the total number of features.

It is clear that the number of potential solutions is exponential, and it is difficult to obtain exact solutions in polynomial time. The main reason is that the required computation grows exponentially with the size of the problem (number of features). Therefore, it is often preferred to look for near-optimal solutions to these problems by using heuristic or metaheuristic algorithms.

3. The Proposed Feature Selection Methods Based on Crowding Distance

In this work, two feature selection algorithms based on crowding distance are proposed. The first one is a filter method while the second is a wrapper method. Both the two algorithms use crowding distance to sort the features. The use of the descending crowding distance as criteria to sort the features is motivated by the following assumption: the most isolated features have the great impacts on the target feature (class) while two closed (or crowded) features A and B have a close impact on the target feature. Therefore, it is preferably to select first the most isolated features than the most crowded features. The crowding distance is adapted as follows: [14]

First of all, the features are sorted according to all the samples (instances) Sm where a sample Sm plays the role of an objective function in multiobjective optimization problems. The vectors of sorted indices I_m are found. The crowding distance CD for each feature is computed using the following equation:

$$CD\big(I_m(i)\big) = \sum_{m=1}^{M} \frac{Sm\big(I_m(i+1)\big) - Sm\big(I_m(i-1)\big)}{S_{m,max} - S_{m,min}} \tag{3}$$

where $I_m(i)$ is the i-th index from the m-th vector of indices, $S_{m,\max}$ and $S_{m,\min}$ are the maximal and minimal values of the m-th Sample data, respectively. The value CD_m for the two extreme features is set to infinity. Geometrically, the crowding distance is the average side length of the cuboid defined by features surrounding a particular feature (see Fig. 1). The less crowded features with a great value of *CD* are the preferred features.

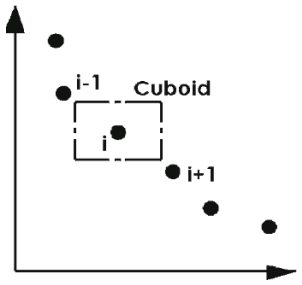

Fig.1. Crowding distance.

3.1. Filter Algorithm Based Cowding Distance

Figure 2 displays the outline of the proposed filter algorithm based crowding distance called FCA for Filter Crowded Algorithm. After reading the dataset, the algorithm computes the crowding distance for every feature, then the obtained distances are sorted in descending order. Finally, the selected features are ranked in the top features of the ordered feature vector. As we can observe, the proposed filter approach is simple and doesn't use any statistical measures.

Input: X=dataset with m samples and n features, k: number of selected features

1. CD=Compute the crowding distance of n features
2. Orderedfeatures =sort (CD, descending)
3. Feats=Select the k first features in the ordered features
4. Compute the accuracy of the reduced dataset X(Feats) by using a certain classifier

Fig.2. The pseudo code of the Filter crowded features.

3.2. Wrapped Algorithm Based Crowding Distance

The wrapped algorithm called WCA (Wrapped Crowded Algorithm) is based on a greedy sequential method where at each step one feature is added to the selected features. If the accuracy of the classifier is enhanced then we keep this feature otherwise it is discarded. The choice of the feature to be added is given by the order of the features computed by the crowding distance. At each step, the fitness of the current solution is computed by a given classifier. Moreover, we can add other termination criteria to stop the algorithm early, like an accuracy threshold. The outline of the proposed algorithm is given in the figure 3.

Input: Read X=dataset with m samples and n features
1. CD=Compute the crowding distance of n features
2. Orderedfeatures=sort (CD, descending)

3. **For** i=1 to n

Selectedfeatures= Selectedfeatures U { Orderedfeatures (i)}
Fitness=Compute the accuracy of the selected features by objective function F
If Fitness >Bestaccuracy
Update the bestaccuracy
Save the selectedfeatures;
Else
Delete the last added features:
Selectedfeatures / { Orderedfeatures (i)}
end

Fig.3. The pseudo code of the wrapped crowded features.

3.3. Hybrid Pearson-crowding Filter

The crowding distance doesn't incorporate any information about the target feature, so for some datasets, the accuracy is lowered. To overcome this shortcoming, the Pearson correlation is integration in this measure in order to enhance the accuracy of the filter method. The new measure distance is an aggregation distance computed by the following equation:

$$\text{distance} = \alpha * \text{crowding\_distance} + \beta * \text{pearason\_coefficient} \quad (4)$$

where α and β are two weights, in this version we have set them to 1, and 1/n respectively, n is the number of features in a given dataset (chosen after an empirical study).

The resulting algorithm called Hybrid Crowded Pearson Algorithm (HCPA) takes advantages of both crowding distance and correlation coefficient which helps to improve the accuracy of the classification.

3.4. Algorithms Complexity

The complexities of the proposed algorithms FCA, WCA, and HCPA are as follows:

$$O(FCA) = O(\text{crowding distance}) + O(\text{sorting algorithm}) = O(m * n * \log(n)) + O(n * \log(n)) \quad (5)$$

$$O(WCA) = O(\text{crowding distance}) + O(\text{sorting algorithm}) + O(\text{selection of k features}) = O(m * n * \log(n)) + O(n * \log(n)) + O(n * O(\text{classifier})) \quad (6)$$

$$O(HPCA) = O(\text{crowding distance}) + O(\text{Pearson coefficient}) + O(\text{sorting algorithm}) = O(m * n * \log(n)) + O(n) + O(n * \log(n)) \quad (7)$$

As we can see, the standard version is the fastest, and the wrapper method is the slowest because it calls iteratively the classifier model used to evaluate the current solution.

4. Implementation and Results

The proposed feature selection algorithms were implemented under MATLAB R2016a environment, and all experiments were carried out on a Windows 10 64-bit computer with an Intel i3 (2.3 GHz) processor and 4 GB RAM. To assess the performance of the developed methods, several popular datasets were used. The details about the used datasets are described in table 1; the mean accuracy of each dataset is also given. In the first phase of this experiment, we have evaluated the filter and wrapped algorithms based on crowding distance. In the second phase, we have evaluated the hybrid method to show the effectiveness of the combination between crowding distance and correlation coefficient. Due to the randomness of the k-fold cross-validation, for one dataset, each algorithm is executed 30 times, and the best, mean, std, and worst results are reported. For all the algorithms, the multiclass SVM classifier [15] was used with kfold=5 in the first experiment and kfold=10 for large datasets.

Table 1. Experimental Datasets.

Dataset	Data Type	# features	# samples	Accuracy all features
Ionosphere	Radar Data	34	351	94.76
Breast	Breast cancer Data	30	569	92.51
Heart	Heart disease	44	267	79.55
Sonar	Signals Rocks vs Mines	60	208	85.42
Ovarian	Ovarian cancer Data	4000	216	96,26
Colon	Colon cancer Data	2000	62	83.00
MLL	Mixed-lineage leukemias Microarray	12533	72	95.14
Semeion	Semeion Handwritten Digit Data	256	1593	97.86
Ovarian-micro	Ovarian Microarray	15154	253	98.13
coiltemp	Columbia Object Image Library	1024	1440	99.96
Zoo	Zoo dataset	101	16	95.94

Table 2 shows the experimental results found by our filter algorithm called Filter Crowded Algorithm (FCA). Moreover, our algorithm is compared to the most popular filter algorithms: Pearson Correlation Coefficient [3], Relief Feature [4] and Variance Feature Selection [2].

Table 2. The experimental results of filter algorithms for feature selection.

dataset	# features selected	Filter Crowded Algorithm				Pearson Correlation Coefficient				Relief Feature				Variance Feature Selection			
		mean	std	worst	best	mean	std	worst	best	mean	Std	worst	Best	mean	std	worst	best
Ionosphere	10	94.13	0.73	92.31	95.16	92.19	0.68	90.60	93.16	95.23	0.42	94.29	**96.02**	89.07	0.52	88.03	90.04
Breast	10	92.52	0.39	91.92	93.31	92.37	0.33	91.74	93.32	92.58	0.37	92.09	93.49	92.53	0.24	92.09	93.15
Heart	10	78.25	0.60	76.76	79.40	79.60	1.10	76.78	81.64	79.05	1.40	76.04	82.01	79.62	1.50	76.42	**82.77**
Sonar	10	71.53	1.97	67.31	74.58	75.83	1.29	73.54	78.34	83.45	1.52	80.73	**86.05**	77.62	1.68	73.53	81.73
Ovarian	150	94.52	0.75	92.12	95.43	89.64	1.15	86.56	91.66	94.62	0.85	93.04	96.30	95.26	0.69	93.99	**96.74**
Colon	150	82.38	1.90	77.56	85.64	83.02	2.29	78.97	**87.31**	81.39	1.57	77.56	83.97	81.40	1.43	78.97	84.10

From table 2, we observe that there is no great difference between our algorithm and the other algorithms. Our algorithm gives a weak accuracy only in sonar dataset and it gives accuracy close to the other algorithms in the reminder datasets. Indeed, the Wilcoxon test confirms that there is no significant difference between the proposed algorithm and the other algorithms at level 0.05.

Table 3 reports the experimental results of the wrapped version. As we can see, the proposed algorithm is able to find a small set of features with higher accuracy. In all the datasets, the best accuracy is greater than 84% and all the results are better than those of the filter algorithms (table 2). The importance of the crowding ordering is clear in the case of ovarian and colon datasets where the number of the selected features is 27 over 4000 features for ovarian dataset and 23 over 2000 features for colon datasets.

In the last experiment, the performance of the hybrid crowded Pearson algorithm (HCPA) on hard datasets including three Microarray datasets. The microarray datasets are real challenges for researchers due to their large number of features. In this experiment, the effectiveness of the hybrid version is clear in most datasets. It ranks first in five tests however, Pearson correlation filter ranks first in three datasets, and FCA is successful in two tests (table 4). This experiment proves that the combination of crowding distance and correlation coefficient in a single measure leads to a more powerful filter method.

Table 3. Statistical results of the wrapped crowded features algorithm.

	# features for best accuracy	**mean**	**std**	**worst**	**best**
Ionosphere	16/34	94.87	0.80	93.15	**96.30**
Breast	8/30	93.15	0.44	92.09	**94.91**
Heart	12/44	81.64	1.33	79.40	**84.63**
Sonar	16/60	86.32	1.48	83.17	**89.38**
Ovarian	27/4000	93.23	2.39	87.04	**96.78**
Colon	23/2000	85.45	2.70	79.36	**91.92**

Table 4. Statistical results of the hybrid algorithm.

		Hybrid crowded pearson				Pearson correlation				FCA			
dataset	features	mean	std	min	max	mean	std	min	max	mean	std	min	max
semeion	20	**99.05**	0.07	98.87	99.12	95.49	0.16	95.10	95.73	89.25	0.17	88.83	89.52
colon	150	**83.73**	1.29	79.76	85.95	82.36	1.82	77.14	85.71	83.08	1.23	79.52	85.71
Ovarianmicro	50	99.44	0.22	98.82	99.62	**99.97**	0.10	99.60	100	79.02	1.30	75.46	80.65
MLL	300	93.10	0.78	91.43	94.64	**94.42**	0.63	92.86	95.89	93.15	0.85	91.61	94.82
Breast cancer	10	92.31	0.32	91.56	92.80	92.47	0.26	91.92	92.80	**92.55**	0.25	92.09	93.14
Ovarian cancer	150	95.63	0.55	94.46	96.77	90.94	1.10	88.90	92.62	**95.80**	0.64	94.42	97.16
coiltemp	100	**99.11**	0.14	98.75	99.31	98.42	0.21	97.92	98.75	86.56	0.31	85.97	87.50
sonar	10	**78.03**	1.09	76.02	80.31	75.37	0.81	73.02	77.00	73.06	1.62	68.19	76.93
zoo	10	92.82	0.70	90.00	94.00	**93.37**	0.47	93.00	94.18	92.83	0.83	90.00	94.09
heart	10	**79.02**	0.81	76.78	80.90	78.86	0.98	76.79	80.91	78.35	0.68	76.41	79.42

5. Conclusions

In this paper, three algorithms for feature selection are presented. The main feature of the proposed algorithm is the use of the crowding distance to order the features from the most important to the less important. The first algorithm is a filter method, the second is wrapped method and the last is a hybrid method between crowding distance and Pearson coefficient. The experimental results show the effectiveness of the proposed algorithms compared to most popular feature selections algorithms. The filter and the hybrid techniques offer better balance between accuracy and speed. In the future, we work to improve the proposed algorithms. The wrapped version can be improved by introducing more specific operators. On the other hand, the hybrid method could be improved by using other statistics metrics like mutual information or variance.

References

[1] Kuhn, M., & Johnson, K. (2013). An introduction to feature selection. In Applied predictive modeling (pp. 487-519). Springer, New York, NY.

[2] Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. Neurocomputing, 300, 70-79.

[3] Sabilla, S. I., Sarno, R., & Triyana, K. (2019). Optimizing threshold using pearson correlation for selecting features of electronic nose signals. Int. J. Intell. Eng. Syst, 12(6), 81-90.

[4] Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. Journal of biomedical informatics, 85, 189-203.

[5] Somol, P., Pudil, P., & Kittler, J. (2004). Fast branch & bound algorithms for optimal feature selection. IEEE Transactions on pattern analysis and machine intelligence, 26(7), 900-912.

[6] Liu, L., Kang, J., Yu, J., & Wang, Z. (2005, October). A comparative study on unsupervised feature selection methods for text clustering. In 2005 International Conference on Natural Language Processing and Knowledge Engineering (pp. 597-601). IEEE.
[7] Sharma, M., & Kaur, P. (2021). A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem. Archives of Computational Methods in Engineering, 28, 1103-1127.
[8] Ogutu, J. O., Schulz-Streeck, T., & Piepho, H. P. (2012, December). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. In BMC proceedings (Vol. 6, No. 2, pp. 1-6). BioMed Central.
[9] Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE transactions on evolutionary computation, 6(2), 182-197.
[10] Bouzoubia, S., Layeb, A., & Chikhi, S. (2014). A multi-objective chemical reaction optimisation algorithm for multi-objective travelling salesman problem. International Journal of Innovative Computing and Applications, 6(2), 87-101.
[11] Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., ... & Cohen, I. (2009). Pearson correlation coefficient. Noise reduction in speech processing, 1-4.
[12] Ghanem, K., & Layeb, A. (2021). Feature Selection and Knapsack Problem Resolution Based on a Discrete Backtracking Optimization Algorithm. International Journal of Applied Evolutionary Computation (IJAEC), 12(2), 1-15.
[13] Hamla, H., & Ghanem, K. (2021). Comparative Study of Embedded Feature Selection Methods on Microarray Data. In Artificial Intelligence Applications and Innovations: 17th IFIP WG 12.5 International Conference, AIAI 2021, Hersonissos, Crete, Greece, June 25–27, 2021, Proceedings 17 (pp. 69-77). Springer International Publishing.
[14] Layeb, A. (2021). Two novel feature selection algorithms based on crowding distance. arXiv preprint arXiv:2105.05212.
[15] Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189-215.

Authors' Profiles

Abdesslem Layeb is professor in the department of computer science at the University of Constantine. I received my PhD degree in computer science from the University of Constantine, Algeria. I'm interested in the combinatorial optimization methods and their applications to solve several problems from different domains like transportation problems, Bioinformatics and other academic problems.
ORCID ID: https://orcid.org/0000-0002-6553-8253

How to cite this paper: Abdesslem Layeb, "Novel Feature Selection Algorithms Based on Crowding Distance and Pearson Correlation Coefficient", International Journal of Intelligent Systems and Applications(IJISA), Vol.15, No.2, pp.37-42, 2023. DOI:10.5815/ijisa.2023.02.04