# A New Way for Combining Filter Feature Selection Methods

**Waad Bouaguel and Mohamed Limam**

**Abstract** This study investigates the issue of obtaining stable ranking from the fusion of the result of multiple filtering methods. Rank aggregation is the process of performing multiple runs of feature selection and then aggregating the results into a final ranked list. However, a fundamental question of is how to aggregate the individual results into a single robust ranked feature list. There are a number of available methods, ranging from simple to complex. Hence we present a new rank aggregation approach. The proposed approach is composed of two stages: in the first we evaluate he similarity and stability of single filtering methods then, in the second we aggregate the results of the stable ones. The obtained results on the Australian and German credit datasets using support vector machine and decision tree confirms that ensemble feature ranking have a major impact in the performance improvement.

**Keywords** Feature selection · Credit scoring

## 1 Introduction

The principal purpose of the dimensionality reduction process is, given a high dimensional dataset $\mathbf{x}_i$, $i \rightsquigarrow x_i = (x_i^1, x_i^2, \ldots, x_i^d)$ that describes a target variable $Y_i$ using $d$ features, to find the smallest pertinent set of features $\mathbf{X} = (X^1, X^2, \ldots, X^d)$, which represent the target variable as all the original set of features do [1, 2]. The process of feature selection in one of the most important task in the pre-analyse that not only consists in finding a reduced set of feature but also the choice of appropriate set based on their pertinence to the study [3].

W. Bouaguel (✉) · M. Limam
LARODEC, ISG, University of Tunis, Tunis, Tunisia
e-mail: bouaguelwaad@mailpost.tn

M. Limam
e-mail: mohamed.limam@isg.rnu.tn

M. Limam
Dhofar University, Salalah, Oman



A. Nagar et al. (eds.), *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*, Smart Innovation, Systems and Technologies 43, DOI 10.1007/978-81-322-2538-6_43

We consider the case of a financial dataset containing data about credit applicants. The class feature, is represented by the solvability level of an applicant, who can be credit worthy or not credit worthy. The class of solvability is assigned to each credit applicant by the credit mangers of the bank. Each customer is represented through a set of features that represent his current credit situation, the past credit history, duration of credits in months, behavior repayment of other loans, value of savings or stocks, stability in the employment, etc. [4].

In general, classification methods use collected information of each credit applicant to classify new ones. Feature reduction, in this context, is employed to find the minimal set of feature which can be used to represent the class of a new credit applicant.

Irrelevant and redundant features decrease the classification performance because they are usually mixed with the relevant ones witch confuse classification algorithms [5], feature selection is useful in this case in order to construct robust predictive models. Many feature selection methods are proposed in literature such as filter and wrapper methods [5]. Filter methods study the fundamental properties of each feature independently of the classifier [6]. In opposite to filters, wrappers use the classifier accuracy to evaluate the feature subsets [7]. Wrappers are the most accurate, but in this case accuracy comes with an exorbitant cost caused by repetitive evaluation [5].

According to [8] filter methods outperforms wrapper methods in many cases. However with the huge number of classical filter methods is difficult to identify which of the filter criteria would provide the best output for the experiments [9, 10]. Then, the best approach is to perform rank aggregation.

According to [11] rank aggregation improve the robustness of the individual feature selection methods such that optimal subsets can be reached [12]. Rank aggregation have many merits. However, an important number of different rank aggregation methods have been proposed in the literature witch make the choice difficult [11].

Thus, this paper discusses the major issues of filter approach and presents a new approach based on rank aggregation. Evaluations on a credit scoring problem demonstrate that the new feature selection approach is more robust and efficient.

This paper is organized as follows. Section 1 briefly reviews majors issues of filter feature selection and give the most famous filtering techniques. Section 3 describes our proposed approach. Experimental investigations and results on two datasets are given in Sect. 4. Finally, Sect. 5 provides conclusions.

## 2 Filter Framework and Rank Aggregation

In this section we will try to give an overall description of some of the most popular univariate filtering methods. Filter methods have many advantages but the most obvious ones are their computational efficiency and feasibility [13]. This advantage allows decision makers to create a complete picture of the available information by examining the data from different angles of various filtering approaches without

**Table 1** Popular filter feature selection methods

| Distance | Dependence | Information |
|---|---|---|
| **Euclidean distance**: Measure the root of square differences between features of a pair of instance | **Pearson**: Measure of linear dependence between two variables | **Mutual Information**: Measure the amount of information shared between two features |
| **Relief**: Measure the relevance of features according to how well their values separate the instances of the same and different classes that are near each other | **Chi-squared**: Measure the statistical independence of two events | **Information Gain**: Information gain but normalized by the entropy of an attribute. Addresses the problem of overestimating the features with multiple values |

increasing the computational complexity and that what makes filter methods extremely effective in practice.

As discussed before, filters select relevant features regardless of the classification algorithm using a independent evaluation function. According to Dash [14], these independent evaluation functions may be grouped into four categories: distance, information, dependence and consistency, where the first three are the most used [15]. Each category have its own specificity and may have large number of filtering methods. Table 1 give the list of the most popular filter feature selection methods.

Filtering methods or further rankers choose one of the independent function discussed before to rank feature according to their relevance to the class label by giving a score to each feature. In general a high score indicates the presence of a pertinent and relevant feature and all features are sorted in decreasing order according to their scores [16]. Many ranks are available in the literature, making the choice difficult for a particular task [8]. According to [17, 18] there is no single best feature ranking method and can not chose the most appropriate filter, unless we evaluate all existing rankers, which is impossible to realise in most domains.

According to [18] rank aggregation which is an ensemble approach for filter feature selection that combine the results of different rankers, might produce a better and more stable ranking than individual rankings. Hence, in this work we investigate a new method combining several rankings.

# 3 Proposed Approach for Combining Filter Feature Selection Methods

## 3.1 *First Ranking for Single Filters and Stability Control*

Many studies show that the stability of an ensemble feature selection model is a curtail topic that influences the final result and the future classification [19]. According to [12] a stable feature selection method is preferred over unstable one in the construction of the final ensemble. Hence we begin by reducing statistical variations of each individual filter in order to retain just the stable ones. According to [19] the stability of each ranker may be quantified by its sensitivity to the variations in the training set. Hence we quantify the stability of each filter by the ranks they give to each feature on several iterations. Each filter was run 10 times for each dataset. In each run a feature ranking is obtained for each filtering method.

According to [19] the stability of a ranker can be measured using a measure of similarity for the ranking representation. Then, we use the Spearman footrule distance [20] as a simple way to compare two ordered lists. Spearman distance between two lists is defined as the sum overall the absolute differences between the ranks of all features from both lists [7]. According to [7, 20], As the Spearman value decreases as the similarity between the two lists increases. Then, the final stability score is the mean of similarity over all the lists for the evaluated filter. Once the final stability score is computed for each filter, we choose the stable ones for the next step. Hence we compare stability score with a threshold of 80 %. If the stability of a filter is less or equal to 80 % the selected filter is conserved for aggregation else it is considered as no stable and eliminated. Figure 1 illustrates the process of choosing the stable filters.

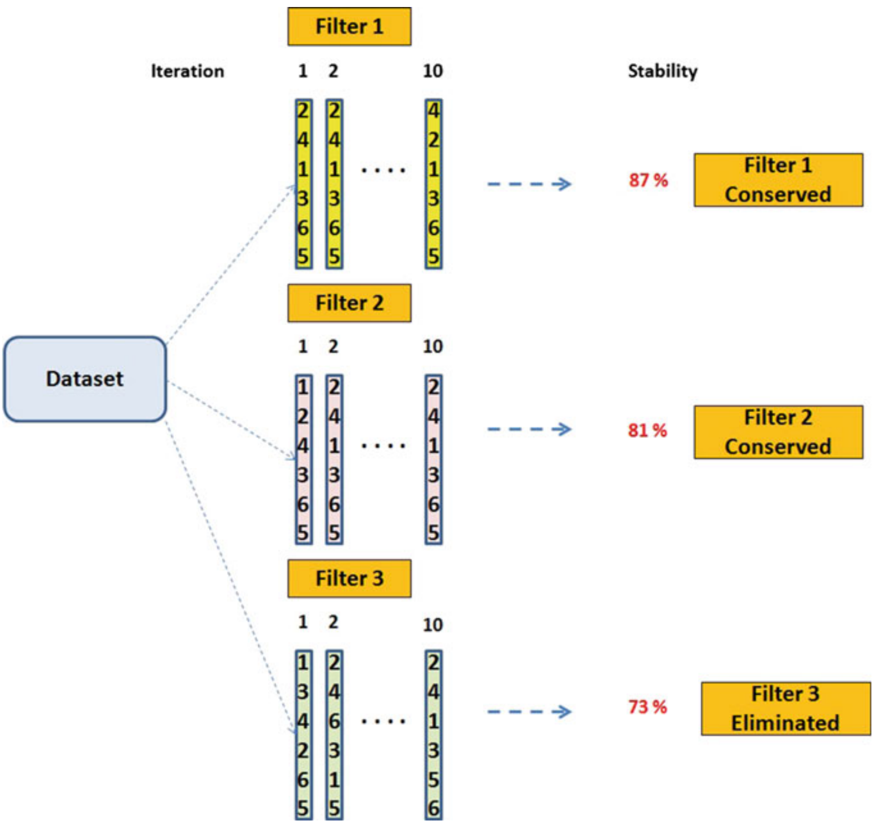


**Fig. 1** The process of choosing the stable filters


**Fig. 2** The fusion process

### *3.2 Merging Different Filter Methods*

Once the most stable filter are selected by the previous stage we move to their combination to provide a more robust result where the issue of selecting the appropriate filter is alleviated to some level [12].

Several rankers are independently applied to find different ranked lists of the same size. Then, these lists of features are merged by selecting feature by feature from each list, starting from the feature on the top of each list and so on [21]. Figure 2 illustrates the fusion process for an example of three filters.

## 4 Experimental Investigations

### *4.1 Datasets Description and Performance Measures*

The adopted herein datasets used for evaluation are the Australian and German credit datasets from the UCI repository of machine learning, available on these links:

**Table 2** Australian dataset: comparison between the new filter method and the other feature selection methods

| | P | R | FM | ROC area |
|---|---|---|---|---|
| *DT* | | | | |
| Relief | 0.682 | 0.813 | 0.742 | 0.575 |
| MI | 0.829 | 0.770 | 0.801 | 0.542 |
| $\chi^2$ | 0.832 | 0.761 | 0.804 | 0.580 |
| Mean | 0.829 | 0.792 | 0.810 | 0.600 |
| Median | 0.831 | 0.789 | 0.808 | 0.613 |
| New approach | **0.850** | **0.854** | **0.852** | **0.662** |
| *SVM* | | | | |
| Relief | 0.695 | 0.798 | 0.743 | 0.602 |
| MI | 0.831 | 0.770 | 0.800 | 0.611 |
| $\chi^2$ | 0.818 | 0.835 | 0.827 | 0.590 |
| Mean | 0.823 | 0.843 | 0.828 | 0.620 |
| Median | 0.821 | **0.845** | 0.832 | 0.621 |
| New approach | **0.845** | 0.821 | **0.833** | **0.798** |

- http://www.cse.ust.hk/~qyang/221/Assignments/German.
- https://archive.ics.uci.edu/ml/datasets/Statlog+(Australian+Credit+Approval).

To implement and test our approach we use four individual filtering techniques from Weka software namely: relief, $\chi^2$ mutual information, and correlation [22]. Each filter is performed 10 time and the first three filters are retained as the most stable one with a stability score over 80 %. The aggregation of these retained filters is performed with Spearman distances.

The obtained results by our proposed approach are compared to two well known rank aggregation techniques: mean, median [7, 23] and also compared to the results given by the individual feature selection methods. Decision trees DT and support vector machine SVM are used as classifiers to evaluate the obtained feature subsets.

The performance of our proposed method is evaluated using three performance measure from the information retrieval field [7]: precision (P), recall (R) and F-measure (FM) and results are presented in Tables 2 and 3.

We investigate the recall results for the set of feature selection methods. For the Australian dataset the best recalls are achieved by our approach for the DT and with the median aggregation for SVM classifier. For the German dataset Table 3 shows that the highest recall is achieved in two times by the new aggregation method with DT and SVM classifiers. We remark from Tables 2 and 3 that the results of aggregation techniques outperform the results of individual feature selection methods, this confirm our hypotheses that aggregation bring more robustness and stability to individual classifiers results. From Tables 2 and 3 we notice that for precision and F-measure the proposed approach always archives the best results.

Graphical tools can be also used as an evaluation criterion instead of a scalar criterion. In this section we use the area under the ROC curve to evaluate the effect of selected features on classification models. Hence, the best combination of

**Table 3** German dataset: comparison between the new filter method and the other feature selection methods

| | P | R | FM | ROC area |
|---|---|---|---|---|
| *Decision tree* | | | | |
| Relief | 0.682 | 0.555 | 0.669 | 0.631 |
| MI | 0.516 | 0.534 | 0.525 | 0.621 |
| $\chi^2$ | 0.737 | 0.477 | 0.579 | 0.600 |
| Mean | 0.750 | 0.542 | 0.612 | 0.682 |
| Median | 0.750 | 0.545 | 0.613 | 0.727 |
| New approach | **0.782** | **0.601** | **0.689** | **0.725** |
| *Support vector machine* | | | | |
| Relief | 0.517 | 0.511 | 0.514 | 0.692 |
| MI | 0.603 | 0.534 | 0.566 | 0.701 |
| $\chi^2$ | 0.705 | 0.489 | 0.577 | 0.622 |
| Mean | 0.766 | 0.552 | 0.627 | 0.780 |
| Median | 0.756 | 0.560 | 0.643 | 0.781 |
| New approach | **0.823** | **0.812** | **0.817** | **0.812** |

features is the one that gives the highest area under the ROC curve will be considered as the most suitable for the classification task.

If we look in ROC area results' we notice from Tables 2 and 3 that proposed approach achieves the highest values with German dataset for both DT and SVM and respectively with the Australian dataset.

## 5 Conclusion

A new approach for rank aggregation in a feature selection context was presented in this study. We tried to implement a robust model for ranking based on ensemble feature selection. In a first part we investigated the stability of filtering methods then we conduct an aggregation on the most stable one. Results on two credit datasets show a remarkable improvement when using our new rank aggregation method compared to the individual rankers and other competitive aggregation methods taken as input. To simplify our work we used a simple similarity criterion, it would be better to study other similarity measure to compute the degree of stability of filtering methods.

## References

1. Ben Brahim, A., Bouaguel, W., Limam, M.: Feature selection aggregation versus classifiers aggregation for several data dimensionalities. In: Proceedings of the International Conference on Control, Engineering & Information Technology (CEIT13) (2013)
2. Ben brahim, A., Bouaguel, W., Limam, M.: Combining feature selection and data classification using ensemble approaches: application to cancer diagnosis and credit scoring.

In: Francisr, T. (ed.) Case Studies in Intelligent Computing: Achievements and Trendss. CRC Press, Boca Raton (2013)
3. Fernandez, G.: Statistical data mining using SAS applications. In; Chapman & Hall/Crc: Data Mining and Knowledge Discovery. Taylor and Francis, Boca Raton (2010)
4. Forman, G.: BNS feature scaling: an improved representation over TF-IDF for SVM text classification. In: Proceedings of the 17th ACM Conference on Information and Knowledge Mining, pp. 263–270. ACM, New York, NY, USA (2008)
5. Rodriguez, I., Huerta, R., Elkan, C., Cruz, C.S.: Quadratic programming feature selection. J. Mach. Learn. Res. **11**(4), 1491–1516 (2010)
6. Saeys, Y., Inza, I.N., Larrañaga, P.: A review of feature selection techniques in bioinformatics. Bioinformatics **23**(19), 2507–2517 (2007)
7. Bouaguel, W., Bel Mufti, G., Limam, M.: A new feature selection technique applied to credit scoring data using a rank aggregation approach based on: optimization, genetic algorithm and similarity. In: Francisr, T. (ed.) Knowledge Discovery & Data Mining (KDDM) for Economic Development: Applications, Strategies and Techniques. CRC Press, Chicago (2014)
8. Wu, O., Zuo, H., Zhu, M., Hu, W., Gao, J., Wang, H.: Rank aggregation based text feature selection. In: Proceedings of the Web Intelligence, pp. 165–172. (2009)
9. Wang, C.M., Huang, W.F.: Evolutionary-based feature selection approaches with new criteria for data mining: a case study of credit approval data. Expert Syst. Appl. **36**(3), 5900–5908 (2009)
10. Bouaguel, W., Bel Mufti, G.: An improvement direction for filter selection techniques using information theory measures and quadratic optimization. Int. J. Adv. Res. Artif. Intell. **1**(5), 7–11 (2012)
11. Dittman, D.J., Khoshgoftaar, T.M., Wald, R., Napolitano, A.: Classification performance of rank aggregation techniques for ensemble gene selection. In: Boonthum-Denecke, C., Youngblood, G.M. (eds.) Proceedings of the International Conference of the Florida Artificial Intelligence Research Society (FLAIRS), AAAI Press, Coconut Grove (2013)
12. Saeys, Y., Abeel, T., Peer, Y.: Robust feature selection using ensemble feature selection techniques. In: Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases—Part II. ECML PKDD '08, pp. 313–325. Springer, Berlin, Heidelberg (2008)
13. Molina, L.C., Belanche, L., Nebot, A.: Feature selection algorithms: a survey and experimental evaluation. In: Proceedings of the IEEE International Conference on Data Mining, pp. 306–313. IEEE Computer Society (2002)
14. Dash, M., Liu, H.: Consistency-based search in feature selection. Artif. Intell. **151**(1–2), 155–176 (2003)
15. Krishnaiah, P., Kanal, L.: Preface. In: Krishnaiah, P., Kanal, L. (eds.) Classification Pattern Recognition and Reduction of Dimensionality. Handbook of Statistics, vol. 2, pp. v–ix. Elsevier (1982)
16. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**(9), 1157–1182 (2003)
17. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics. Springer New York Inc, New York (2001)
18. Prati, R.C.: Combining feature ranking algorithms through rank aggregation. In: The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. Brisbane, Australia, 10–15 June 2012
19. Kalousis, A., Prados, J., Hilario, M.: Stability of feature selection algorithms: a study on high-dimensional spaces. Knowl. Inf. Syst. **12**(1), 95–116 (2007)
20. Pihur, V., Datta, S., Datta, S.: RankAggreg, an R package for weighted rank aggregation. BMC Bioinform. **10**(1), 62–72 (2009)

21. Mak, M.W., Kung, S.Y.: Fusion of feature selection methods for pairwise scoring svm. Neurocomputing **71**(16–18), 3104–3113 (2008)
22. Bouckaert, R.R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., Scuse, D.: Weka manual (3.7.1) (2009)
23. Kolde, R., Laur, S., Adler, P., Vilo, J.: Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics **28**(4), 573–580 (2012)