Application of Multi-Objective Optimization to Feature Selection for a Difficult Data Classification Task

Joanna Grzyb[0000−0003−0698−6993], Mariusz Topolski[0000−0002−5213−6845] and Michał Woźniak[0000−0003−0146−4205]

Department of Systems and Computer Networks,
Faculty of Electronics, Wrocław University of Science and Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
{joanna.grzyb,mariusz.topolski,michal.wozniak}@pwr.edu.pl

Abstract. Many different decision problems require taking a compromise between the various goals we want to achieve into account. A specific group of features often decides the state of a given object. An example of such a task is the feature selection that allows increasing the decision's quality while minimizing the cost of features or the total budget. The work's main purpose is to compare feature selection methods such as the classical approach, the one-objective optimization, and the multi-objective optimization. The article proposes a feature selection algorithm using the Genetic Algorithm with various criteria, i.e., the cost and accuracy. In this way, the optimal Pareto points for the nonlinear problem of multi-criteria optimization were obtained. These points constitute a compromise between two conflicting objectives. By carrying out various experiments on various base classifiers, it has been shown that the proposed approach can be used in the task of optimizing difficult data.

Keywords: *multi-objective optimization*, *feature selection*, *cost-sensitive*, *classification*

1 Introduction

The development of modern information technologies leading to increasingly faster digitization of every aspect of human life makes information systems necessary to process more data. With technological progress, the problem of acquiring and storing large amounts of training data for machine learning models disappeared. As a consequence, the volume of features describing a given object was increased. This, in turn, caused a deterioration in the quality of the classification.

The reason why more information does not mean better classification is the so-called *Curse of Dimensionality* also known as *small n, large p* [1] and it was first described by Richard Bellman [2]. As dimensions are added to the feature set, the distances between specific points are constantly increasing. Additionally, the number of objects needed for correct generalization is increasing. Nonparametric classifiers, such as Neural Networks or those that use radial basis

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

functions, where the number of objects required for valid generalization grows exponentially, are worse at this problem [3].

The *Hughes phenomenon* arises from the the *Curse of Dimensionality* [4]. For a fixed number of samples, the recognition accuracy may increase with increasing features but decreases if the number of attributes exceeds a certain optimal value. This is due to the distance between the samples and the noise in the data or irrelevant features.

The above-mentioned *Curse of Dimensionality* makes it challenging to classify objects in a task when a specific decision is made based on the analysis of multiple criteria. In some complex problems, we use not one but many criteria to make the final decision. There is no general concept of optimality in such tasks because there are various objective criteria that are often contradictory: the optimal solution for one criterion may differ from the optimal one for another criterion. Therefore, there is no single optimality criterion for measuring the solution quality for many such real problems. One popular approach to deal with such a task is the Pareto optimization [5]. The Pareto optimality concept, named in honor of the Italian scientist Vilfredo Pareto, is a compromise of many goals of solving some complex and challenging problem [6].

Finding or approximating a set of non-dominated solutions and choosing among them is the main topic of multi-criteria optimization and multi-criteria decision making. There are many methods used in the multi-criteria optimization task, including Simplex models, methods based on graphs, trees, two-phase Simplex, etc. Decision trees are compared in [7]. The lower cost limit obtains similar or better quality for some data sets because it uses a greedy approach that does not guarantee globally optimal solutions.

Modern research in multi-criteria optimization uses methods that, apart from improving classification accuracy, also affect supervised classifiers' generalization ability. One such approach uses an unsupervised grouping procedure based on ascending Growing Hierarchical Self-Organising Maps (GHSOM) to select features [8]. In [9], Jiang et al. proposed a wrapper framework for Test-Cost-Sensitive Feature Selection (TCSFS) where the difference between the accuracy and the total cost create the evaluation function. The two-objective optimization problem was transformed into one-objective. However, the method reaches good accuracy and a low total test cost. The same objectives but in the form of the two-objective optimization case were considered in [10] as the Two-Archive Multi-Objective Artificial Bee Colony Algorithm (TMABC-FS). For solving the Cost-Sensitive Feature Selection (CSFS) problem, this method is a good alternative.

For comparison with above-mentioned techniques, other methods of feature selection are used in tasks with one criterion, such as PCA (*Principal Component Analysis*), ICA (*Independent Component Analysis*) [11], LDA (*Linear Discriminant Analysis*) applied to the linear combination [12] and PCA modification in the form of CCPCA (*Centroid Class Principal Component Analysis*) [13] or GPCA (*Gradient Component Analysis*) [14]. Also some statistical methods are applied to select features: *ANOVA* [15] i.e. the analysis of the variance and *Pearson's correlation coefficient* [16, 17] which are better at a single criterion task.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

Here it is also worth mentioning *wrapper* methods which make selection based on the analysis of the results of a specific classifier [18, 15].

Multi-criteria optimization is widely used in many fields and is gaining increasing interest. It is a promising field of research. Important applications of the multi-criteria optimization include the minimization of various types of error rates in the machine learning (false positive, false negative) [19, 20], the optimization of delivery costs and inventory costs in logistics [21], the optimization of building design in terms of health, energy efficiency and cost criteria [22]. Other applications can be found in medicine. For example, when looking for new therapeutic drugs, we maximize a drug's potency while minimizing the cost of synthesis and undesirable side effects [23].

One should notice that considering the cost of feature acquisition, we encounter additional feature selection limitations. In some cases, if an additional maximum budget is given, some of the features may not be taken into account. Therefore, features' cost should be considered as an additional factor introducing difficulties in the data for the feature selection task. Considering the importance of cost-sensitive data classification, we decided to tackle this problem in our paper. It is possible to obtain a lower total test cost and the accuracy comparable (not worse) to other methods. Hence our contributions are:

- The method proposal of using the two-objective optimization *NSGAII* algorithm and ranking its solutions according to the accuracy, the total cost, or *PROMETHEE*.
- The experimental evaluation of the proposed method with the classical approach and the one-objective optimization *Genetic Algorithm*.

2 Methods

Ikram and Cherukuri mentioned that *Chi-square* is the best method for multi-class problems [24] and a few of our chosen data sets are multi-class, we used this method as the feature selection technique. The *Chi-square* test statistic applied to the *Select K-best* function choose K-features from the data sets, which are the most relevant to the classification process. This is the classical approach, and we refer to it as *FS* (Feature Selection).

Many papers address the problem of feature selection using one-objective optimization such as *Genetic Algorithms* (*GA*) [25]. *GA* searches a population, and through the iteration, it evaluates and uses genetic operators (selection, mutation, crossover) to finds the best solutions [26]. In our experiment, we used *GA* with two different objective functions. Firstly, the maximum accuracy score has been applied and we indicate it as *GA-a*.

$$\text{maximize } g_1 = accuracy \tag{1}$$

Secondly, we aggregated the accuracy score and selected features' total cost to obtain a cost-sensitive classifier marked as *GA-ac*.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

$$\text{maximize } g_2 = \frac{accuracy}{cost} \tag{2}$$

However, using only one-objective can be insufficient, and even the aggregating process is not enough. Hence, the better approach is to used the multi-objective optimization where each criterion is considered separately [25]. From several algorithms we chose *NSGAII* (Non-dominated Sorting Genetic Algorithm II) [27] – the updated multi-objective version of *GA*. It was used in the feature selection problem [28]. In our experiment, *NSGAII* has been applied with two fitness functions, and each of them is treated independently. The accuracy has to be maximized and the total cost - minimized.

$$\begin{cases} \text{maximize } f_1 = accuracy \\ \text{minimize } f_2 = cost \end{cases} \tag{3}$$

The diagram of the general genetic algorithm is in Figure 1. The binary representation is an example of 6 features of the *liver* data set. The bit string is a vector of features called an individual or a solution, where 1 means that the feature was selected by the algorithm and 0 - not selected. In the beginning, random sampling is performed, so the initial set of solutions is created. Then, the binary tournament random mating selection is used. N-individuals are selected in each tournament, where $n = 2$ in our case. Individuals are compared with each other, and the winner is taken to the next generation population. It is a simple and efficient solution ensuring diversity [29]. Next, two genetic operators are applied to produce new offsprings: the binary point crossover and the bit-flip mutation. The selection is used to choose significant solutions to create the population and genetic operators explore the search space. As shown in Figure 1, the crossover swaps the part of the bit string, and the mutation replaces the bit with the opposite value. The search is over when the algorithm reaches the population size.

The *NSGAII* algorithm returns the non-dominated set of solutions (called the Pareto front) from which one solution must be chosen to contain a subset of the best features. These features are used to learn the classifier and obtain the best performance and the lowest total cost. However, the best features are not the same when there are different expectations. Sometimes the total test cost or the accuracy is more important, but sometimes there is a need to have good two criteria. Therefore, we applied three approaches to ranking solutions. Firstly, the criterion based on the maximum accuracy was chosen as the *NSGA-a* method. Secondly, the solution with the minimum cost was selected as *NSGA-c*. Lastly, the *PROMETHEE II* [30] approach was implemented as *NSGA-p*. This approach is a pair-wise comparison that returns the ranking of solutions from the best to the worst. It requires criteria weights and the preference function. Then, based on outranking flows (positive and negative), the complete ranking is obtained [31].

We compare all methods described above and believe the multi-objective optimization approach can achieve a more inexpensive total test cost without an

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

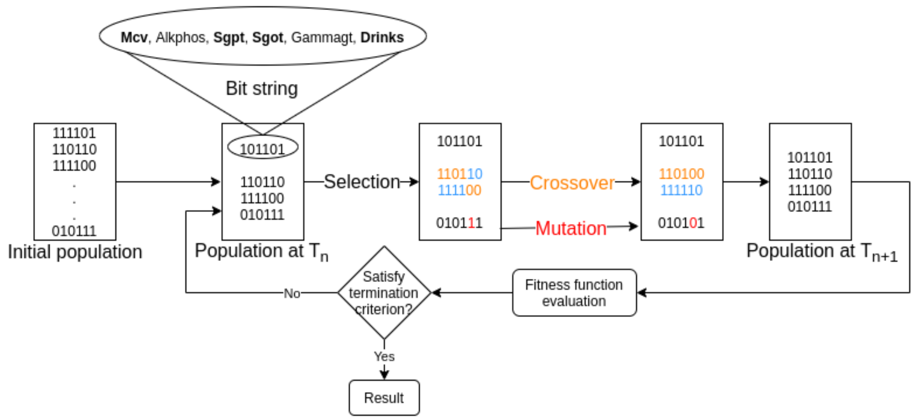


Fig. 1: Diagram of the genetic algorithm and operators

accuracy drop compared to other methods. To prove our hypothesis, we conduct the experimental evaluation in the following section.

3 Experimental evaluation

Our research tries to find answers to two research questions:

RQ1: How do feature selection methods work for each classifier?
RQ2: Can feature selection method based on multi-objective optimization outperform classical methods based on one-objective optimization?

3.1 Setup

Data sets with the corresponding features' cost are obtained from the UCI Machine Learning Repository [32]. All of them are medical data sets where the total cost of tests is important, so classification is not a trivial task. The information about the number of examples, attributes, and classes is in Table 1. The aim of the first data set *heart-disease* is to predict if a patient has heart disease. The *hepatitis* data set contains information about patients with Hepatitis disease and decision 1 or 2 (die or survive). The *liver-disorders* data set has the smallest number of features. Based on them, the decision of a person who suffers from alcoholism is made. The *pima-indians-diabetes* contains only female medical data from the Pima Indians group (Native Americans), in which class says if a person has diabetes or not. The last data set *thyroid-disease* is the biggest one and it has many features. It contains three classes that decide if an individual is normal or suffers from hyperthyroidism or hypothyroidism (1, 2, or 3). We consider only these data sets in our experiment since we do not know where the feature cost is.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

Data sets	**Number of examples**	**Number of attributes**	**Number of classes**
heart-disease	303	13	4
hepatitis	155	19	2
liver-disorders	345	6	2
pima-indians-diabetes	768	8	2
thyroid-disease	7200	21	3

Table 1: Data sets

The project is implemented in the Python programming language and it is available in the GitHub repository [1] along with results from the experiment. A few libraries were used: Pymoo [33], Matplotlib [34], Pandas [35], Numpy [36] and scikit-learn [37]. From the last one we used following classifiers with the default parameters:

- Decision Tree Classifier - *CART*
- Support Vector Machines - *SVM*
- Gaussian Naive Bayes - *GNB*
- K-Nearest Neighbors Classifier - *kNN*

Before experiments, all data sets must be preprocessed. First, missing values were replaced with the most frequent ones. Then, data and features' cost were normalized. The number of examples is relatively small, so Repeated Stratified K-Fold (5 splits x 2 repeats) cross-validation was used to avoid overfitting. Lastly, mechanisms of feature selection described in Section 2 were used. The population size of all genetic algorithms was set to 100.

After the preprocessing experiments were run, and based on the accuracy score, which measures the performance, the comparison of methods is made.

3.2 Experiment

While multi-objective optimization is used, the algorithm finds many solutions, and it returns only the Pareto set. Figure 2 shows three non-dominated solutions in the form of black dots, each of which has corresponding values of the accuracy score and the total cost of the selected features. The greater the accuracy, the higher the cost, so choosing which solution to choose is crucial. Hence, we applied three different criteria described in Section 2 to select the best solution. In the *PROMETHEE* approach, the weight 0.4 was used for the accuracy and the weight 0.6 for the cost.

The experiment compares six feature selection methods along with four classifiers tested on five data sets. The number of features changes from 1 to the maximum features in the data set. 24 micro charts are showed for each data set, in which an orange line represents the accuracy and a blue line - the total cost.

[1] https://github.com/joannagrzyb/moofs

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

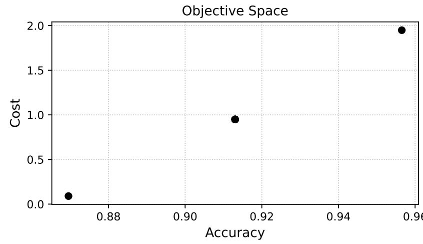


Fig. 2: Pareto front, data set: *hepatitis*

The cost of each feature has been normalized to a value from 0 to 1, so the total cost is the sum of these values, not the original ones from the UCI and it is on the second y-axis on the right. The y-axis on the left contains the accuracy score, and the x-axis - number of features. All approaches to the feature selection problem with the abbreviation coming up in figures are presented in Table 2.

Methods	**Objectives**	**Criteria**	**Abbr**
Select K-best (Chi-square) Feature Selection	-	-	FS
Genetic Algorithm	max. accuracy	-	GA-a
	max. (accuracy/cost)	-	Ga-ac
Non-dominated Sorting Genetic Algorithm II	max. accuracy min. cost	max. accuracy	NSGA-a
		min. cost	NSGA-c
		PROMETHEE	NSGA-p

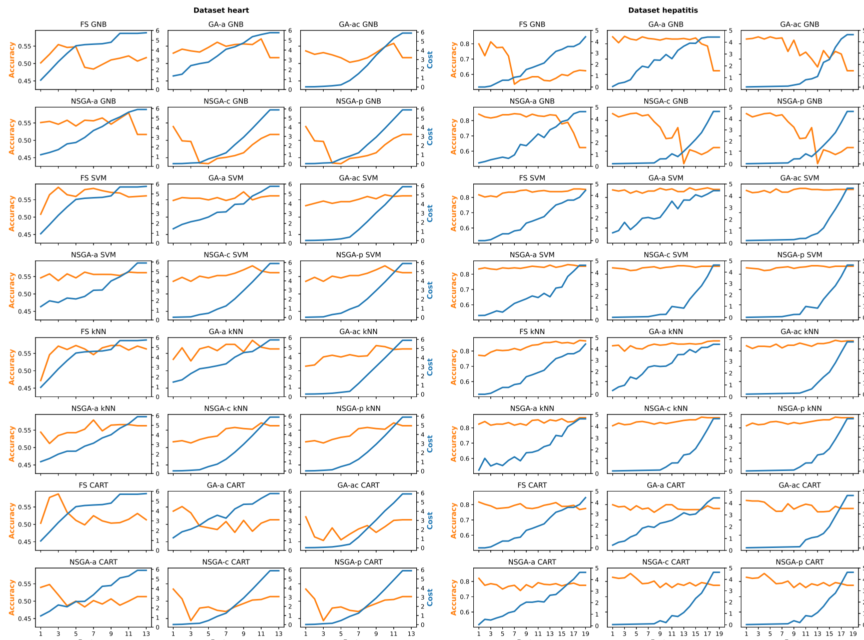
Table 2: Methods' abbreviation

Figure 3a shows results for the *heart* data set. For optimization methods, the total cost has a shape similar to the exponential function, unlike the classical approach *FS*, in which the total cost grows very fast. For *SVM* and *kNN*, the accuracy is stable for all methods, so it is cost-effective to choose the smaller number of features because the accuracy is almost the same, but the total cost is much smaller for *GA-ac*, *NSGA-c*, *NSGA-p*. The optimal number of features, in this case, is 4 or 5. For remaining classifiers, the tendency is not the same and too many features lead to a deterioration of the classification's quality.

Figure 3b shows results for the *hepatitis* data set. It can be seen that using all features to learn the classifier is not always the best idea. This data set contains 19 features, and for *GNB* many features disturb in good classification. We can obtain the same accuracy level for other classifiers but a much smaller total cost using only 5 features and optimization methods with the cost criterion. Figure 4 shows values of the accuracy and the total cost for all tested approaches using *SVM* classifier in the *hepatitis*. The accuracy is very stable among all methods and through a different number of selected features. Furthermore, as we observed

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

earlier, the total cost is much smaller for three methods *GA-ac*, *NSGA-c* and *NSGA-p*.



(a) *heart* data set (b) *hepatitis* data set

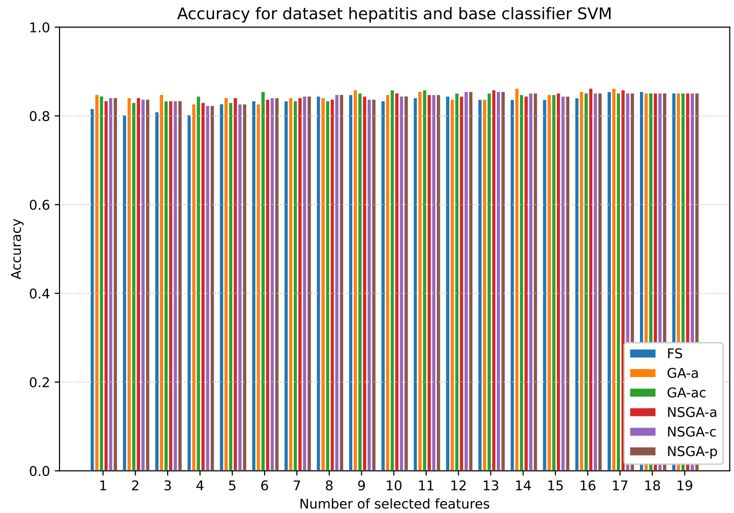
Fig. 3: Accuracy and cost

The *liver* data set (Figure 5a) has the smallest number of features, so the difference between classical approaches and optimization ones is not very large, especially for the cost, which is almost linear. However, the accuracy for non-classical methods has a bigger value than for *FS*. The optimal feature number is 3 for *SVM* and *kNN* where methods prefer the accuracy (*GA-a*, *GA-ac*, *NSGA-a*) and also they keep the low total cost. Even if you do not need the cost-sensitive classifier, it is better not to use *FS* in that case because it has a smaller performance.

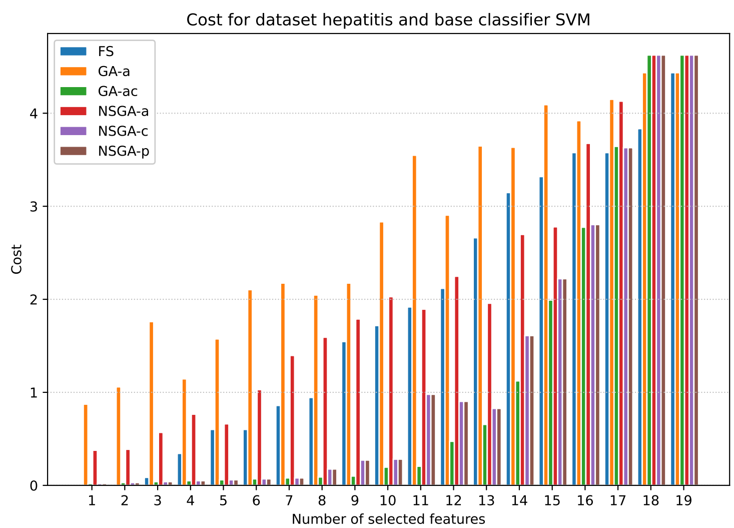
In the *pima* data set in Figure 5b the cost is the smallest for methods with the cost criterion (*GA-ac*, *NSGA-c*, *NSGA-p*), but in the same time they obtain the smallest accuracy. As in the previous case, the optimal number of features is 3. In that point, the *GA-a* and *NSGA-a* for *GNB* achieve the highest accuracy over 75% and the lowest cost under 1.

As in the previous data set, the *thyroid* data set (Figure 6) has the similar cost shape of *GA-ac*, *NSGA-c* and *NSGA-p* methods and they achieve very small total cost. For them, along with SVM and kNN, the optimal number of selected

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8



(a) Accuracy

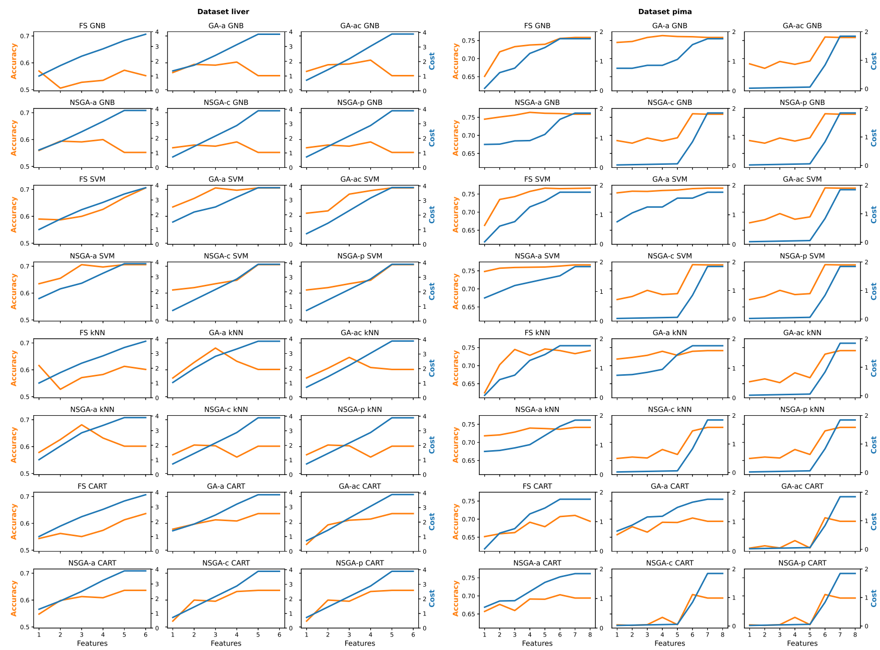


(b) Cost

Fig. 4: Bar charts for data set *hepatitis* and *SVM* classifier

features is 9 with the cost close to 0 and the accuracy around 95%. Overall, for this data set, the classification quality is much bigger than in other data sets because thyroid has a few thousands times more instances.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8



(a) *liver* data set (b) *pima* data set

Fig. 5: Accuracy and cost

3.3 Lessons learned

After conducted experiments, we can answer the research question posed at the beginning of this section:

RQ1: How do feature selection methods work for each classifier?
GNB is not recommended as the accuracy is usually smaller than for other classifiers and it has greater discrepancy among different number of features. Otherwise *SVM* and *kNN* gives the stable accuracy score among all feature selection techniques in *heart-disease*, *hepatitis* and *thyroid* data sets. Unlike these data, the accuracy varies in each technique through different number of selected features from 5% to 10% in *liver-disorders* and *pima-indians-diabetes* data sets. *CART* is similar to other methods.

RQ2: Can feature selection method based on multi-objective optimization outperform classical methods based on one-objective optimization ?
Suppose there is a need to have a good performance classification and a low total cost. In that case, it is worth considering the multi-objective optimization algorithms to select the best features. The quality of the classification depends on the data. However, it can be useful in the medical environment where some tests can be costly. Thanks to this approach, a person can select only tests that give good results during the classification of the disease, and at the same time, the cost will be low.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

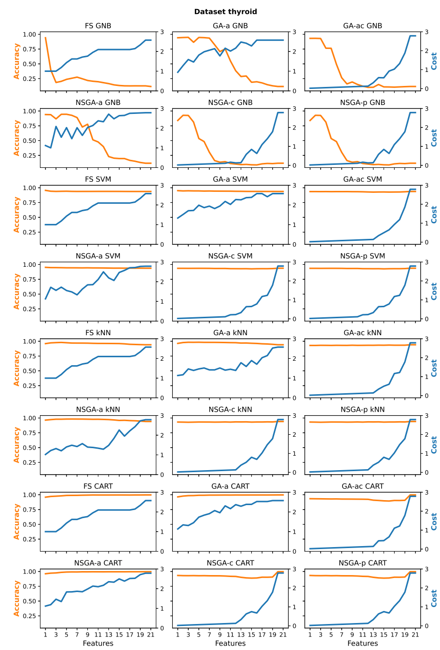


Fig. 6: Accuracy and cost for data set thyroid

4 Conclusions

The work's purpose was to propose a new method and compare a few different feature selection approaches in the classification data with feature cost. The selection of K-best features with the Chi-square statistic, *Genetic Algorithm* in the form of the one-objective and multi-objective optimization were applied to classifiers, such as Gaussian Naive Bayes, Support Vector Machines, K-Nearest Neighbors and Decision Tree Classifier. In our method, we expected a much lower cost and accuracy close to other methods.

Based on the conducted experiments, we gather two criteria to consider if the aim is to achieve good classification and the low total cost of features. Depending on the data set, the one-objective method or the two-objective method should be used to obtain better results than the classical approach, so our proposition is quite effective and feasible.

Further research directions include increasing the number of data sets and taking into account various optimization criteria. It will be interesting to have data with more features and patterns used to learn and test the developed methods. Although *Chi-square* is one of the best feature selection methods, other methods can be used, which may be more effective with different optimization criteria.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

Acknowledgements

This work was supported by the Polish National Science Centre under the grant No. 2019/35/B/ST6/04442.

References

1. Gersende Fort and Sophie Lambert-Lacroix. Classification using partial least squares with penalized logistic regression. *Bioinformatics*, 21(7):1104–1111, 2005.
2. Richard E Bellman. *Adaptive control processes: a guided tour*, volume 2045. Princeton university press, 2015.
3. Luis O Jimenez and David A Landgrebe. Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2653–2667, 1999.
4. Gordon Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, 14(1):55–63, 1968.
5. Allen Klinger. Letter to the editor—improper solutions of the vector maximum problem. *Operations Research*, 15(3):570–572, 1967.
6. Nodari Vakhania and Frank Werner. A brief look at multi-criteria problems: Multi-threshold optimization versus pareto-optimization. In *Multi-criteria Optimization-Pareto-optimal and Related Principles*. IntechOpen, 2020.
7. Wojciech Penar and Michal Wozniak. Cost-sensitive methods of constructing hierarchical classifiers. *Expert Systems*, 27(3):146–155, 2010.
8. Emiro De la Hoz, Eduardo De La Hoz, Andrés Ortiz, Julio Ortega, and Antonio Martínez-Álvarez. Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps. *Knowledge-Based Systems*, 71:322–338, 2014.
9. Liangxiao Jiang, Ganggang Kong, and Chaoqun Li. Wrapper framework for test-cost-sensitive feature selection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.
10. Yong Zhang, Shi Cheng, Yuhui Shi, Dun-wei Gong, and Xinchao Zhao. Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm. *Expert Systems with Applications*, 137:46–58, 2019.
11. Kailash J Karande and Rajashree N Badage. Facial feature extraction using independent component analysis. In *Annual Int'l Conference on Intelligent Computing, Computer Science & Information Systems (ICCSIS-16)*, pages 28–29, 2016.
12. Riddhi A Vyas and SM Shah. Comparision of pca and lda techniques for face recognition feature based extraction with accuracy enhancement. *International Research Journal of Engineering and Technology (IRJET)*, 4(6):3332–3336, 2017.
13. Mariusz Topolski. The modified principal component analysis feature extraction method for the task of diagnosing chronic lymphocytic leukemia type b-cll. *Journal of Universal Computer Science*, 26(6):734–746, 2020.
14. Mariusz Topolski. Application of the stochastic gradient method in the construction of the main components of pca in the task diagnosis of multiple sclerosis in children. In *International Conference on Computational Science*, pages 35–44. Springer, 2020.
15. Andrea Bommert, Xudong Sun, Bernd Bischl, Jörg Rahnenführer, and Michel Lang. Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational Statistics & Data Analysis*, 143:106839, 2020.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

16. Jie Cai, Jiawei Luo, Shulin Wang, and Sheng Yang. Feature selection in machine learning: A new perspective. *Neurocomputing*, 300:70–79, 2018.
17. Diah Risqiwati, Adhi Dharma Wibawa, Evi Septiana Pane, Wardah Rahmatul Islamiyah, Agnes Estuning Tyas, and Mauridhi Hery Purnomo. Feature selection for eeg-based fatigue analysis using pearson correlation. In *2020 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pages 164–169. IEEE, 2020.
18. Beatriz Remeseiro and Veronica Bolon-Canedo. A review of feature selection methods in medical applications. *Computers in biology and medicine*, 112:103375, 2019.
19. Iryna Yevseyeva, Vitor Basto-Fernandes, David Ruano-OrdáS, and José R MéNdez. Optimising anti-spam filters with evolutionary algorithms. *Expert systems with applications*, 40(10):4010–4021, 2013.
20. Pu Wang, Michael Emmerich, Rui Li, Ke Tang, Thomas Bäck, and Xin Yao. Convex hull-based multiobjective genetic programming for maximizing receiver operating characteristic performance. *IEEE Transactions on Evolutionary Computation*, 19(2):188–200, 2014.
21. Martin Josef Geiger and Marc Sevaux. The biobjective inventory routing problem–problem solution and decision support. In *International Conference on Network Optimization*, pages 365–378. Springer, 2011.
22. Christina J Hopfe, Michael TM Emmerich, Robert Marijt, and Jan Hensen. Robust multi-criteria design optimisation in building design. *Proceedings of building simulation and optimization, Loughborough, UK*, pages 118–125, 2012.
23. Susanne Rosenthal and Markus Borschbach. Design perspectives of an evolutionary process for multi-objective molecular optimization. In *International Conference on Evolutionary Multi-Criterion Optimization*, pages 529–544. Springer, 2017.
24. Ikram Sumaiya Thaseen and Cherukuri Aswani Kumar. Intrusion detection model using fusion of chi-square feature selection and multi class svm. *Journal of King Saud University-Computer and Information Sciences*, 29(4):462–472, 2017.
25. Grandchamp Enguerran, Mohamed Abadi, and Olivier Alata. An hybrid method for feature selection based on multiobjective optimization and mutual information. *Journal of Informatics and Mathematical Sciences*, 7(1):21–48, 2015.
26. Laura Emmanuella A dos S Santana and Anne M de Paula Canuto. Filter-based optimization techniques for selection of feature subsets in ensemble systems. *Expert Systems with Applications*, 41(4):1622–1631, 2014.
27. Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.
28. Utkarsh Singh and Shyam Narain Singh. Optimal feature selection via nsga-ii for power quality disturbances classification. *IEEE Transactions on Industrial Informatics*, 14(7):2994–3002, 2017.
29. Noraini Mohd Razali, John Geraghty, et al. Genetic algorithm performance with different selection strategies in solving tsp. In *Proceedings of the world congress on engineering*, volume 2, pages 1–6. International Association of Engineers Hong Kong, 2011.
30. Gang Kou, Pei Yang, Yi Peng, Feng Xiao, Yang Chen, and Fawaz E Alsaadi. Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing*, 86:105836, 2020.
31. Majid Behzadian, Reza Baradaran Kazemzadeh, Amir Albadvi, and Mohammad Aghdasi. Promethee: A comprehensive literature review on methodologies and applications. *European journal of Operational research*, 200(1):198–215, 2010.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8

32. Moshe Lichman et al. Uci machine learning repository, 2013.
33. J. Blank and K. Deb. Pymoo: Multi-objective optimization in python. *IEEE Access*, 8:89497–89509, 2020.
34. J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
35. Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.
36. Travis E Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.
37. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_8