A Decision Rule Based Approach to Generational Feature Selection

Wiesław Paja(✉)

Faculty of Mathematics and Natural Sciences, University of Rzeszów,
1 Pigonia Street, 35-310 Rzeszów, Poland
wpaja@ur.edu.pl

Abstract. The increase of dimensionality of data is a target for many existing feature selection methods with respect to efficiency and effectiveness. In this paper, the all relevant feature selection method based on information gathered using generational feature selection is described. The successive generations of feature subset were defined using *Rule Quality Importance* algorithm and next the subset of most important features was eliminated from the primary dataset. This process was executed until the most relevant feature has got importance value on the level equal to importance of the random, shadow feature. The proposed approach was also tested on well-known artificial Madelon dataset and the results confirm its efficiency. Thus, the conclusion is that the identified features are relevant but not all weakly relevant features were discovered.

Keywords: Decision rules · Generational feature selection
Feature ranking · All relevant feature selection
Dimensionality reduction

1 Introduction

In the era of high-dimensional datasets some methods that provide possibilities to effective analysis of such kind of data are crucial. These kind of methods are applied to improve performance in terms of speed, predictive power and simplicity of the model [1]. Moreover, they are used to visualize the data for model selection, to reduce dimensionality and remove noise of themselves. For this reason a feature selection methods are required. The feature selection is a process that able to choose an optimal subset of features according to a certain criterion. In this way, irrelevant data (features) could be removed from the original set. It able to increase the predictive accuracy of developed learning models, reduce the cost of the data. It also can improve the learning efficiency, such as reducing storage requirements and computational cost. Finally, it could be observed the reduction of the complexity of the resulting model description, improving the understanding of the data and the model. Feature selection can be considered as a search problem, where each state of the search space corresponds to

P. Perner (Ed.): ICDM 2018, LNAI 10933, pp. 230–239, 2018.
https://doi.org/10.1007/978-3-319-95786-9_17

a concrete subset of features selected [2]. Thus, two main directions of search can be identified: *Sequential Forward Generation* (*SFG*) and *Sequential Backward Generation* (*SBG*). According to *SFG* approach searching process starts with an empty set of features *F*. As the search starts, features are added into *F* according to some criterion that distinguish the relevant feature from the others. Features set *F* grows until it reaches some stopping criteria. It can be a threshold for the number of relevant features m or simply the generation of all possible subsets in brute force mode. In turn, the *SFG* approach searching process starts with a full set of features and, iteratively, they are removed one at a time. Here, the criterion must point out the least relevant feature. Finally, the subset contains only a unique feature, which is considered to be the most informative of the whole set. As in the previous case, different stopping criteria can be used. The two other direction of search can be also recognized: the first one *Bidirectional Generation* (*BG*) begins the search in both directions, performing *SFG* and *SBG* concurrently. They stop in two cases: when one search finds the best subset comprised of m features before it reaches the exact middle, or both searches achieve the middle of the search space. It takes advantage of both *SFG* and *SBG*. The second method *Random Generation* (*RG*) starts the search in a random direction. The choice of adding or removing a features is a random decision. It tries to avoid the stagnation into a local optima by not following a fixed way for subset generation. Unlike *SFG* or *SBG*, the size of the subset of features cannot be stipulated.

During the selection process two main goals are identified [3]:

- *Minimal Optimal Feature Selection* (MOSF), where the goal is to discover the minimal subset of features with the best classification quality;
- *All Relevant Feature Selection* (ARFS), where the main goal is to discover all informative features, even that with minimal relevance [4,5];

Here, presented approach is focused on the second type of FS. Motivation for this methodology was *Recursive Feature Elimination* (*RFE*) algorithm in which by application of external estimator specific weights values are assigned to each features [6,7]. This procedure is repeated recursively, and in each step, features whose weights are the smallest are removed from the currently investigated set of features. It works until the expected set of features to select is eventually obtained. In *RFE* approach a number of feature to select should be initially defined. In turn, in presented approach, the number of feature is unknown and to distinguish between relevant and irrelevant features the *contrast variable* concept [5] has been applied. Contrast does not carry information on the decision attribute by design but it is added to the system in order to discern relevant and irrelevant (shadow) attributes. Contrast values are obtained from the original features by random permutation of values through the objects of the analyzed dataset. The use of contrast variables was for the first time proposed by Stoppiglia [8] and then by Tuv [9]. In this way, the goal of proposed methodology is to simplify and improve feature selection process by relevant feature selection during recursive generation of decision rules. The hypothesis is that by removing subsets of relevant features in each step gradually all-relevant feature subset could be indicated.

2 Methods and Algorithms

In this section, theoretical description of applied methods and algorithms is shortly presented. Mainly, the *Generational Feature Selection* approach and connected methods for contrast feature generation and decision rule quality based importance estimation are described.

2.1 The Decision Rule Quality Based Importance Algorithm

During research the *DRQualityImp* algorithm [10,11] is used to define ranking values for each investigated feature. This algorithm (see *Algorithm 1*) is based on the presence of different feature in decision rule set generated from dataset. Thus, the ranking measure *RQI* (*Rule Quality based Importance*) for attribute a could be defined.

$$RQI(a) = \sum_{j=1}^{k} q_j \cdot \omega(a) \tag{1}$$

where: k is a number of rules generated; $\omega(a)$ describes the occurrence of the attribute a in j-th rule and q_j denotes the quality of this rule. The pseudocode for this algorithm is presented below. Here, as an input, a special case of a decision system is considered, when only one decision attribute d which determining classes of objects in the decision system S is distinguished, i.e., $S = (U, A \cup \{d\})$. U is the nonempty finite set of objects known as the universe of discourse and

Algorithm 1. Decision Rule Quality based Importance estimation

```
Input  : S = (U, A ∩ {d}) - a decision system; R – a decision rule set; Q –
         quality of rules r ∈ R, cond(r) – a set of conditionals of the rule r.
Output: RQI – a set of importances of features.
Function ruleQualityImportance(A,R,Q)
    RQI ← ϕ
    for each a ∈ A do
        RQI(a) = 0
        for each r ∈ R do
            if a ⊂ cond(r) then
                RQI(a) = RQI(a) + Q(r)
            end
        end
        RQI ← RQI ∪ RQI(a)
    end
    return RQI in decreasing order
end
```

Algorithm 2. Contrast features generation

```
Input  : S = (U, A ∩ {d}) - a decision system.
Output: A_CONT – a set of contrast features.
Function contrastFeatures(A)
    A_CONT ← φ
    for each a ∈ A do
        a_CONT ← permute(V_a)
        A_CONT ← A_CONT ∪ a_CONT
    end
    return A_CONT
end
```

A is the nonempty finite set of attributes (features). Each attribute $a \in A$ is a function $a : U \rightarrow V_a$, where V_a is the set of values of a.

2.2 The Contrast Features Generation Algorithm

Presented approach applies also contrast features algorithm (see *Algorithm 2*) that is used to establish threshold between relevant and irrelevant features inside the investigated set. Here, the decision system $S = (U, A \cup \{d\})$ is also considered as the input. Contrast (shadow) features set A_{CONT} are generated from original set A by the random permutation of each attribute a through their values space V_a. So, as the output the set of contrast attributes is obtained. It is assumed, that these features are not correlated with decision one, and in this way some kind of artificial noise could be added to original data.

2.3 The Generational Feature Selection Algorithm

The experimental procedure, initially called *Generational Feature Selection*, is presented in the form of pseudocode bellow as the *Algorithm 3*. As an input, the decision system $S = (U, A \cup \{d\})$ is considered. Additionally, machine learning algorithm *MLA* utilized during feature importance estimation has to be introduced. The output is in the form of selected important features subset *FS*. Generally, algorithm iteratively generates learning model thus it could be called *generation*. After the first generation (iteration) selected important features are removed from dataset. The next generation is done based on the remaining data, and so on. During each iteration the contrast features A_{CONT} are generated (*contrastFeatures*) based on original current set A_{CURR} and added to it creating extended set A_{EXT}. Using this extended set the learning model m is developed (*generateModel*) and machine learning algorithm *MLA* is applied. Here, decision rule algorithm is utilized. Then, the set of importances M of each feature in developed learning model m is calculated (*modelMeasure*), and selected ranking algorithm is applied (*rankingAlgorithm*) to obtain ranking L. During research it was *ruleQualityImportance* method (see *Algorithm 1*). Next, contrast feature

Algorithm 3. Generational Feature Selection

```
Input  : S = (U, A ∩ {d}) - a decision system; MLA – an applied machine
         learning algorithm.
Output: FS – a selected feature subset.
Function GFS(S,MLA)
    A_CURR ← A
    FS ← φ
    x=0
    while x=0 do
        A_CONT ← contrastFeatures(A_CURR)
        A_EXT ← A_CURR ∪ A_CONT
        m ← generateModel(S, MLA)
        M ← modelMeasure(m)
        L ← rankingAlgorithm(A_EXT, M)
        maxCFRank ← max(L(a : a ∈ A_CONT))
        FS ← φ
        for each l ∈ L(a : a ∈ A_CURR) do
            if l > maxCFRank then
                FS ← FS ∪ a(l)
            end
            A_CURR ← A_CURR \ FS
            if FS = φ then
                x++
            end
            FS ← A \ A_CURR
        end
    end
    return FS
end
```

with maximal value of ranking (*maxCFRank*) is identified. After that, set of relevant features *FS* which have ranking value l higher than *maxCFRank* is identified. Discovered *FS* set is then removed from the currently investigated set A_{CURR}. If *FS* is empty then iterations stop. Finally, *FS* is defined by removing irrelevant features from the original set A.

3 Results and Conclusions

For illustration of the test of proposed algorithm the well-known in the domain of feature selection Madelon dataset is considered. It is an artificial data set, which was one of the Neural Information Processing Systems challenge problems in 2003 (called NIPS2003) [12]. It contains 2600 objects (2000 of training objects + 600 of validation objects) corresponding to points located in 32 vertices of a 5-dimensional hypercube. Each vertex is randomly assigned to the one of two classes: −1 or +1, and the decision of each object is a class of its vertex.

Objects are characterized by 500 features which were constructed in the following way: 5 of them are randomly jittered coordinates of points, the other 15 attributes are random linear combinations of the first 5. The rest of the data is a uniform random noise. The goal is to select 20 important attributes from the system without false attributes selection. Additionally, the same 10 fold cross validation process was applied during experiments to efficient comparison of different approaches. Here, only detailed results for 6^{th} sample fold of the Madelon dataset using *RQI* method based on the *CN2* rule quality or own rule quality measure with the *WRACC* evaluation function (see Tables 1 and 2) and also for 2^{nd} fold with the *Laplace* evaluation function (see Table 3) are presented. As it was shown for example in Table 1, four iterations (generations) of the algorithm were done. During the first iteration, the classification rules (the first generation of rules) have been built based on all input data. Using the *RQI* method based on *CN2* rule quality measure, the subset of fourteen features is indicated as relevant one (grey cells marked), according to decreased values of the *RQI* calculated from the developed decision rules set. Then, the subset of selected features is removed from the dataset. Next, in the 2^{nd} iteration of algorithm the next one, probably relevant, subset is selected using the *RQI* values calculated from the rules developed on the reduced dataset. The second feature, $f203\_1$, which is the contrast feature defines the threshold for selection of the important subset. This subset (just one feature) is also removed from dataset. Next, in the

Table 1. The sample results of importance of features gathered in 6^{th} fold using *RQI* based on ***CN2* rule quality measure** and the ***WRACC* evaluation function**. Bold names denote truly relevant features, others denote irrelevant features. Name with _1 index denote contrast feature. The grey colored cells denote the feature set found important in given iteration (generation) and which is removed from the data in the next iteration.

1st iteration		2nd iteration		3rd iteration		4th iteration	
feature name	RQI CN2	feature name	RQI CN2	feature name	RQI CN2	feature name	RQI CN2
f339	0.225	**f379**	0.035	f5	0.071	f213_1	0.037
f337	0.176	f203_1	0.021	f190	0.059	f31_1	0.035
f242	0.151	f411	0.021	f191	0.056	f186	0.021
f65	0.108	f315	0.021	f135_1	0.053	f73_1	0.020
f473	0.090	f286	0.021	f357	0.033	f53_1	0.016
f443	0.087	f217	0.020	f366	0.032	f249	0.016
f494	0.085	f458	0.020	f485_1	0.032	f357	0.015
f454	0.077	f154	0.019	f327_1	0.032	f205_1	0.015
f476	0.057	f435_1	0.019	f83_1	0.030	f497	0.015
f49	0.051	f497_1	0.019	f40	0.030	f328_1	0.015
f129	0.050	f282	0.017	f166	0.030		
f456	0.048	f5	0.017	f183_1	0.029		
f106	0.035	f357	0.016	f125_1	0.029		
f319	0.025			...	...		

Table 2. The sample results of importance of features gathered in 6^{th} fold using RQI based on ***own* rule quality measure** and the ***WRACC* evaluation function**. Bold names denote truly relevant features, others denote irrelevant features. Name with _1 index denote contrast feature. The grey colored cells denote the feature set found important in given iteration (generation) and which is removed from the data in the next iteration.

1st iteration		2nd iteration		3rd iteration		4th iteration	
feature name	RQI CN2	feature name	RQI CN2	feature name	RQI CN2	feature name	RQI CN2
f339	2.970	**f379**	1.624	f190	2.742	f31_1	1.237
f337	2.616	**f154**	0.84	f5	2.292	f213_1	1.149
f494	2.002	**f282**	0.775	f432	1.909	f357	0.742
f242	1.996	f203_1	0.73	f435_1	1.909	f328_1	0.742
f65	1.924	f411	0.73	f448	1.786	f53_1	0.713
f473	1.375	f217	0.675	f183_1	1.783	f249	0.713
f443	1.294	f458	0.675	f11_1	1.779	f205_1	0.709
f454	1.119	f315	0.674	f478_1	1.773	f497	0.709
f49	1.095	f286	0.674	...	...	f73_1	0.667
f319	0.857	f435_1	0.649			f186	0.633
f476	0.829	f497_1	0.649				
f106	0.742	f5	0.577				
f129	0.711	f357	0.541				
f456	0.632						

3^{rd} iteration of algorithm the next subset of three probably relevant features is found using the RQI values calculated from the rules constructed on the subsequently reduced dataset. The fourth feature, $f135\_1$, defines the threshold for selection of the next important subset. This subset is therefore removed from dataset. Finally, in the 4^{th} iteration the subset of important features is empty, because the highest value of the RQI measure is reached by contrast feature $f213\_1$. In this way, the algorithm stops, and the subset of 18 features is defined as the relevant one. The truly relevant features in Madelon dataset are written in bold. It could be observed that three attributes: $f5$, $f190$ and $f191$, were also included in the discovered subset. However their relevance is very random and unique what is simply presented in Table 4, where these attributes are only 3, 2 and 1 times selected respectively. They reach >0.5 of the feature removing probability threshold P_{rm} during the 10-fold cross-validation. Similar results are presented in Tables 2 and 3. Proposed threshold P_{rm} for a given feature a is defined below.

$$P_{rm}(a) = \frac{ncross - nsel(a)}{ncross} \quad (2)$$

Here, $ncross$ means the number of the validation folds, in turn $nsel$ means the number of selections of the feature a [13].

Table 3. The sample results of importance of features gathered in 2^{nd} fold using *RQI* based on ***CN2*** and ***own*** **rule quality measure** and the ***Laplace*** **evaluation function**. Bold names denote truly relevant features, others denote irrelevant features. Name with _1 index denote contrast feature. The grey colored cells denote the feature set found important in given iteration (generation) and which is removed from the data in the next iteration.

1st iteration			2nd iteration			3rd iteration			4th iteration			5th iteration		
feat. name	RQI CN2	RQI own	feat. name	RQI CN2	RQI own	feat. name	RQI CN2	RQI own	feat. name	RQI CN2	RQI own	feat. name	RQI CN2	RQI own
f339	37.19	40	f106	11.20	12	f5	5.56	6	f39	5.44	6	f104_1	4.55	5
f476	12.21	13	f443	7.47	8	f393_1	5.09	6	f229_1	4.48	5	f244	3.69	4
f242	12.11	13	f494	6.67	7	f342_1	4.57	5	f84_1	3.77	4	f264	3.69	4
f49	9.40	10	f473	4.72	5	f479_1	4.46	5	f375_1	3.75	4	f411	3.68	4
f454	9.39	10	f229_1	4.55	5	f395_1	4.43	5	f270_1	3.74	4	f269_1	3.68	4
f129	7.26	8	f228	4.51	5	f267_1	3.77	4	f141	3.64	4	f74_1	3.68	4
f337	6.61	7	f229	4.41	5	f291_1	3.76	4	f310	3.64	4	f291_1	3.66	4
f379	6.48	7	f405_1	3.98	5	f54_1	3.71	4	f387	3.64	4	f46_1	3.62	4
f65	4.44	5	f319	3.82	4	f206	3.69	4	f492_1	3.50	4	f357	3.61	4
f174	4.42	5	f282	3.80	4	...	...	...	...	...	...	...	...	...
f29	3.84	4	f5	3.73	4									
f466_1	3.69	4	...	...	...									
...	...	...												

The summary results of the feature selection are collected in Table 4, where they are compared with earlier gathered results of experiments using the decision tree formalism (the *DTLevelImp* column) [11]. As it can be seen, Generational Feature Selection based on decision rules discover about $15 \div 16$ truly relevant features. In turn, the decision tree based approach discovered all twenty relevant features without false positives. They didn't exceed the threshold of removing probability > 0.5.

Initial results are promising, however, it could be identified problem with the unequivocal definition of the threshold used to separate truly relevant feature from the other irrelevant. For example, in case of Madelon dataset, the *f5* feature which is random noise was discovered 2, 3 or 4 times during 10-fold cross-validation (see Table 4), thus their probability estimator for removing is even *0.6*. Similar values of this estimator could be observed for known truly relevant features *f456* or *f154*. These features are weakly relevant and they are difficult to detect. Also, the influence of the process of features values discretization may be crucial for discovering them. The proposed algorithm of generational feature selection seems to be robust and let to find weakly relevant important attributes due to sequential elimination of strongly relevant attributes.

Table 4. The summary results gathered in **10-folds** using *RQI* based on ***CN2*** and ***own* rule quality measure** and the ***WRACC* and *Laplace* evaluation functions**. These results are compared with ***DTLevelImp*** algorithm. Bold names denote truly relevant features, other denotes irrelevant ones. The grey colored cells denote the feature set indicated as relevant. *1 - feature name, 2 - # of selections, 3 - removing probability*

WRACC + CN2 RQI			WRACC + own RQI			Laplace			DTLevelImp		
1	2	3	1	2	3	1	2	3	1	2	3
f49	10	0.0	f49	10	0.0	f49	10	0.0	f29	10	0.0
f65	10	0.0	f65	10	0.0	f65	10	0.0	f49	10	0.0
f129	10	0.0	f129	10	0.0	f129	10	0.0	f65	7	0.3
f242	10	0.0	f242	10	0.0	f242	10	0.0	f106	10	0.0
f337	10	0.0	f337	10	0.0	f337	10	0.0	f129	10	0.0
f339	10	0.0	f339	10	0.0	f339	10	0.0	f154	10	0.0
f443	10	0.0	f443	10	0.0	f443	10	0.0	f242	10	0.0
f454	10	0.0	f454	10	0.0	f454	10	0.0	f282	10	0.0
f473	10	0.0	f473	10	0.0	f473	10	0.0	f319	10	0.0
f476	10	0.0	f476	10	0.0	f476	10	0.0	f337	10	0.0
f106	10	0.0	f106	10	0.0	f494	10	0.0	f339	10	0.0
f494	9	0.1	f494	9	0.1	f106	9	0.1	f379	10	0.0
f319	8	0.2	f379	8	0.2	f379	9	0.1	f434	8	0.2
f379	6	0.4	f319	7	0.3	f29	7	0.3	f443	10	0.0
f282	6	0.4	f282	7	0.3	f452	6	0.4	f452	10	0.0
f456	4	0.6	f456	5	0.5	f154	3	0.7	f454	6	0.4
f154	2	0.8	f154	4	0.6	f405	3	0.7	f456	5	0.5
f29	2	0.8	f29	2	0.8	f5	2	0.8	f473	10	0.0
f286	2	0.8	f286	2	0.8	f8	2	0.8	f476	10	0.0
f452	1	0.9	f5	2	0.8	f178	2	0.8	f494	6	0.4
f434	1	0.9	f190	2	0.8	f206	2	0.8	f5	4	0.6
f5	3	0.7	f452	1	0.9	f229	2	0.8	f177	1	0.9
f362	1	0.9	f434	1	0.9	f282	2	0.8	f359	1	0.9
f227	1	0.9	f362	1	0.9	f286	2	0.8	f414	1	0.9
f190	2	0.8	f227	1	0.9	f39	1	0.9	f85	1	0.9
f264	1	0.9	f264	1	0.9	f43	1	0.9	f256	1	0.9
f357	1	0.9	f357	1	0.9	f52	1	0.9	f112	1	0.9
f191	1	0.9	f432	1	0.9	f103	1	0.9	f286	2	0.8
f229	1	0.9	f266	1	0.9	f153	1	0.9	f216	1	0.9
f194	1	0.9	f287	1	0.9	f174	1	0.9	f292	2	0.8
f12	1	0.9	f236	1	0.9	f201	1	0.9	f343	1	0.9
f174	1	0.9	f269	1	0.9	f246	1	0.9	f74	1	0.9
f489	1	0.9				f279	1	0.9	f148	1	0.9
f244	1	0.9				f284	1	0.9	f472	1	0.9
f41	1	0.9				f304	1	0.9	f203	1	0.9
f350	1	0.9				f306	1	0.9	f211	1	0.9
						f319	1	0.9	f304	1	0.9
						f333	1	0.9	f7	1	0.9
						f334	1	0.9	f440	1	0.9
						f335	1	0.9	f323	1	0.9
						f357	1	0.9	f245	1	0.9
						f358	1	0.9			
						f381	1	0.9			
						f403	1	0.9			
						f411	1	0.9			

Acknowledgments. This work was supported by the Center for Innovation and Transfer of Natural Sciences and Engineering Knowledge at the University of Rzeszów.

References

1. Kuhn, M., Johnson, K.: Applied Predictive Modeling. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-6849-3
2. Liu, H., et al.: Feature selection aspects. In: Liu, H., Motoda, H. (eds.) Feature Selection for Knowledge Discovery and Data Mining. The Springer International Series in Engineering and Computer Science, vol. 454, pp. 43–72. Springer, Boston (1998). https://doi.org/10.1007/978-1-4615-5689-3_3
3. Nilsson, R., Peña, J.M., Björkegren, J., Tegnér, J.: Detecting multivariate differentially expressed genes. BMC Bioinform. **8**, 150 (2007)
4. Paja, W., Wrzesień, M., Niemiec, R., Rudnicki, W.R.: Application of all-relevant feature selection for the failure analysis of parameter-induced simulation crashes in climate models. Geosci. Model Dev. **9**, 1065–1072 (2016)
5. Rudnicki, W.R., Wrzesień, M., Paja, W.: All relevant feature selection methods and applications. In: Stańczyk, U., Jain, L.C. (eds.) Feature Selection for Data and Pattern Recognition. SCI, vol. 584, pp. 11–28. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-45620-0_2
6. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Mach. Learn. **46**, 389–422 (2002)
7. Johannes, M., Brase, J.C., Frohlich, H., Gade, S., Gehrmann, M., Falth, M., Sultmann, H., Beiflbarth, T.: Integration of pathway knowledge into a reweighted recursive feature elimination approach for risk stratification of cancer patients. Bioinformatics **26**(17), 2136–2144 (2010)
8. Stoppiglia, H., Dreyfus, G., Dubois, R., Oussar, Y.: Ranking a random feature for variable and feature selection. J. Mach. Learn. Res. **3**, 1399–1414 (2003)
9. Tuv, E., Borisov, A., Torkkola, K.: Feature selection using ensemble based ranking against artificial contrasts. In: International Symposium on Neural Networks, pp. 2181–2186 (2006)
10. Paja, W.: Feature selection methods based on decision rule and tree models. In: Czarnowski, I., Caballero, A.M., Howlett, R.J., Jain, L.C. (eds.) Intelligent Decision Technologies 2016. SIST, vol. 57, pp. 63–70. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39627-9_6
11. Paja, W.: Generational feature elimination to find all relevant feature subset. In: Czarnowski, I., Howlett, R.J., Jain, L.C. (eds.) IDT 2017. SIST, vol. 72, pp. 140–148. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-59421-7_13
12. Guyon, I., Gunn, S., Ben-Hur, A., Dror, G.: Result analysis of the NIPS 2003: feature selection challenge. In: Advances in Neural Information Processing Systems, vol. 17, pp. 545–552 (2013)
13. Paja, W., Pancerz, K., Grochowalski, P.: Generational feature elimination and some other ranking feature selection methods. In: Stańczyk, U., Zielosko, B., Jain, L.C. (eds.) Advances in Feature Selection for Data and Pattern Recognition. ISRL, vol. 138, pp. 97–112. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67588-6_6