

Discriminant Function Selection in Binary Classification Task

Robert Burduk

Abstract The ensemble selection is one of the important problems in building multiple classifier systems (MCSs). This paper presents dynamic ensemble selection based on the analysis of discriminant functions. The idea of the selection is presented on the basis of binary classification tasks. The paper presents two approaches: one takes into account the normalization of the discrimination functions, and in the second approach, normalization is not performed. The reported results based on the data sets from the UCI repository show that the proposed ensemble selection is a promising method for the development of MCSs.

Keywords Ensemble selection · Multiple classifier system · Binary classification task

1 Introduction

The pattern recognition task is one of the trends in research on machine learning [1]. The classification task can be accomplished by a single classifier or by a team of classifiers. In the literature, the use of multiple classifiers for a decision problem is known as MCSs or an ensemble of classifiers [4, 11, 27]. These methods are popular for their ability to fuse together multiple classification outputs for better accuracy of classification. The building of MCSs consists of three phases: generation, selection, and integration [3]. The final decision which is made in the third phase uses the prediction of the selected classifiers. The output of an individual classifier can be divided into three types [17].

R. Burduk (✉)

Department of Systems and Computer Networks, Wrocław University of Technology, Wybrzeże Wyspińskiego 27, 50-370 Wrocław, Poland
e-mail: robert.burduk@pwr.edu.pl

© Springer International Publishing Switzerland 2016

R. Burduk et al. (eds.), *Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015*, Advances in Intelligent Systems and Computing 403, DOI 10.1007/978-3-319-26227-7_25

265

- The abstract level—classifier ψ assigns the unique label j to a given input x .
- The rank level—in this case for each input x , each classifier produces an integer rank array. Each element within this array corresponds to one of the defined class labels. The array is usually sorted with the label at the top being the first choice.
- The measurement level—the output of a classifier is represented by a measurement value that addresses the degree of assigning the class label to the given output x . An example of such a representation of the output is a posteriori probability returned by Bayes classifier.

Following these three types of outputs of the base classifier, various problems of combination function (third phase) of classifier outputs are considered. The problems studied in [19, 25] belong to the abstract level. The combining outputs for the rank level are presented in [13] and problems studied in [16, 18] belong to the last level. The selection of classifiers is one of the important problems in the creation of EoC [15, 24]. This task is related to the choice of a set of classifiers from all the available pool of classifiers. Formally, if we choose one classifier, it is a classifier selection process. But if we choose a subset of classifiers from the pool, then it is an ensemble selection. In this work, these terms will be used interchangeably. Here you can distinguish between the static or dynamic selection [22, 26]. In the static classifier selection one set of classifiers is selected to create EoC. This EoC is used in the classification of all the objects from the testing set. The main problem in this case is to find a pertinent objective function for selecting the classifiers. One of the best objective functions for the abstract level of classifier outputs is the simple majority voting error [23]. In the dynamic classifier selection for each unknown sample a specific subset of classifiers is selected [5]. It means that we are selecting different EoCs for different objects from the testing set. In this type of classifier selection, the classifier is chosen and assigned to the sample based on different features [28] or different decision regions [7, 14]. In this work we will consider the dynamic ensemble selection. In detail, we propose the new selection method based on the analysis of the discriminant functions in the contents of the binary classification. The paper presents two approaches: one takes into account the normalization of the discrimination functions. In the second approach, normalization is not performed. The text is organized as follows: in Sect. 2 the ensemble of classifiers and combination functions based on the sum method are presented. Section 3 contains the new method for the dynamic ensemble selection based on the analysis of the discriminant functions. Section 4 includes the description of research experiments comparing the suggested algorithms with base classifiers. Finally, the discussion and conclusions from the experiments are presented.

2 Ensemble of Classifiers

Let us consider the binary classification task. It means that we have two class labels $M = \{1, 2\}$. Each pattern is characterized by a feature vector X . The recognition algorithm maps the feature space X to the set of class labels M according to the

general formula:

$$\Psi : X \rightarrow M. \quad (1)$$

Let us assume that K different classifiers $\Psi_1, \Psi_2, \dots, \Psi_K$ are available to solve the classification task. In MCSs these classifiers are called base classifiers. In the binary classification task K is assumed to be an odd number. As a result, for all the classifiers' actions, their K responses are obtained. The output information from all K component classifiers is applied to make the ultimate decision of MCSs. This decision is made based on the predictions of all the base classifiers. One of the possible methods for integrating the output of the base classifier is the sum rule. In this method the score of MCSs is based on the application of the following sums:

$$s_i(x) = \sum_{k=1}^K \hat{p}_k(i|x), \quad i \in M, \quad (2)$$

where $\hat{p}_k(i|x)$ is an estimate of the discrimination functions for class label i returned by classifier k . The final decision of the MCSs is made following the maximum rule:

$$\Psi_S(x) = \arg \max_i s_i(x). \quad (3)$$

In the presented method (3) the discrimination functions obtained from the individual classifiers take an equal part in building MCSs. This is the simplest situation in which we do not need additional information about the testing process of the base classifiers except for the models of these classifiers. One of the possible methods in which weights of the base classifier are used is presented in [2].

2.1 Selection of Discrimination Functions

The classifier selection phase is often criticized due to usually observed high computational cost [3]. Bearing this in mind, we propose the dynamic selection of discrimination functions, which has minor computing requirements. It uses the assumption that the more credible the classifier, the larger the differences in the discrimination function. In the discussed binary classification task this means that during the selection we compare the discriminant functions $|\hat{p}_k(1|x) - \hat{p}_k(2|x)| < \alpha$. The parameter α determines the size of the difference in the discriminant functions. The values are derived from the interval $\alpha \in [0, 1)$. For value $\alpha = 0$ the selection process does not occur. The proposed selection process is performed before the calculation of the coefficients $s_i(x)$. The final decision is made according to the sum rule, and MCSs with the selection is labelled as $\Psi_{S_{ON}}^\alpha$. In MSCs model labeled as $\Psi_{S_N}^\alpha$, in addition the normalization of the discrimination functions is carried out. Normalization is performed for each label class i according to the rule:

$$\hat{p}'_k(i|x) = \frac{\hat{p}_k(i|x) - \min(\hat{p}_1(i|x), \dots, \hat{p}_k(i|x))}{\max(\hat{p}_1(i|x), \dots, \hat{p}_k(i|x)) - \min(\hat{p}_1(i|x), \dots, \hat{p}_k(i|x))}, \quad k \in K. \tag{4}$$

The aim of the experiments presented in the next section was among other things, to compare two proposed methods of selection.

3 Experimental Studies

A series of experiments were carried out to illustrate the quality of classifications. The aim of the experiments was to compare the proposed selection method algorithms with the base classifiers and ensemble classifiers based on the majority voting and sum methods. In the experiential research 11 benchmark data sets were used. Two of them were generated randomly—they are the so-called Banana and Highleyman sets. The other nine benchmark data sets come from the UCI repository [10]. All the data sets concern the binary classification task. The numbers of attributes, examples, and ration in the classes are presented in Table 1. The studies did not include the impact of the feature selection process on the quality of classifications. Therefore, the feature selection process [12, 21] was not performed. The results are obtained via tenfold-cross-validation method. In experiment 7, base classifiers were used. Two of them work according to $k - NN$ rule where the k parameter is equal to 5 or 7. Two base classifiers are used as decision trees algorithms, with the number of branches denoted as 2 and the depth of the precision tree having at most 6 levels. In the decision-making nodes the Gini index or entropy is used. One of the classifiers uses a combination of the decision trees' base models. It is Gradient boosting algorithm. The other two classifiers use Support Vector Machines models. One of them uses Least Squares SVM method and the second Decomposed Quadratic Programming

Table 1 Description of data sets selected for the experiments

Data set	Example	Attribute	Ration (0/1)
Banana	400	2	1.0
Blood	748	5	3.2
Breast cancer wisconsin	699	10	1.9
Highleyman	400	2	1.0
Ionosphere	351	34	1.8
Indian liver patient	583	10	0.4
Mammographic mass	961	6	1.2
Parkinson	197	23	0.3
Pima Indians diabetes	768	8	1.9
Sonar (mines versus rocks)	208	60	0.9
Statlog	690	14	0.8

Table 2 Classification accuracy and mean rank positions for the proposed selection algorithm with normalization produced by the Friedman test

Data set	Ψ_{S_N} with $\alpha =$							
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Banana	0.975	0.975	0.975	0.975	0.977	0.977	0.977	0.977
Blood	0.657	0.660	0.655	0.655	0.660	0.665	0.657	0.647
Cancer	0.947	0.945	0.945	0.947	0.947	0.942	0.942	0.940
Higle	0.955	0.955	0.955	0.952	0.952	0.952	0.952	0.952
Ion	0.967	0.970	0.972	0.970	0.960	0.957	0.962	0.955
Liver	0.602	0.597	0.595	0.622	0.647	0.667	0.660	0.662
Mam	0.625	0.620	0.615	0.632	0.637	0.640	0.632	0.632
Park	0.977	0.977	0.987	0.980	0.985	0.985	0.980	0.982
Pima	0.572	0.575	0.572	0.582	0.592	0.577	0.582	0.577
Sonar	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.980
Statlog	0.912	0.912	0.915	0.915	0.915	0.920	0.922	0.910
Mean rank	4.36	4.36	4.09	3.36	2.45	2.72	3.18	4.63

method. The experiments were performed in an SAS Enterprise Miner environment. Table 2 shows the results of the classification for the proposed classifier selection with normalization of the posteriori probability functions. Additionally, the mean ranks obtained by the Friedman test were presented. The values show that the best value of the parameter α is 0.4 for that classifier selection method. Table 3 shows the results

Table 3 Classification accuracy and mean rank positions for the proposed selection algorithm without normalization produced by the Friedman test

Data set	$\Psi_{S_{ON}}$ with $\alpha =$							
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Banana	0.977	0.977	0.977	0.977	0.977	0.977	0.977	0.977
Blood	0.672	0.672	0.660	0.605	0.647	0.660	0.662	0.572
Cancer	0.942	0.945	0.945	0.945	0.940	0.940	0.937	0.935
Higle	0.952	0.952	0.950	0.952	0.952	0.957	0.957	0.940
Ion	0.962	0.962	0.965	0.962	0.957	0.960	0.950	0.960
Liver	0.675	0.675	0.655	0.662	0.665	0.665	0.665	0.617
Mam	0.642	0.637	0.632	0.632	0.632	0.635	0.625	0.625
Park	0.987	0.990	0.990	0.992	0.990	0.987	0.987	0.982
Pima	0.600	0.602	0.597	0.605	0.595	0.587	0.585	0.580
Sonar	0.987	0.987	0.985	0.985	0.985	0.980	0.980	0.977
Statlog	0.910	0.910	0.907	0.907	0.910	0.912	0.910	0.907
Mean rank	2.18	1.63	3.63	3.09	3.81	3.63	4.54	6.81

Table 4 Classification accuracy and mean rank positions for the base classifiers (ψ_1, \dots, ψ_7) and MCSs algorithms produced by the Friedman test

Data set	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7	ψ_{MV}	ψ_{Sum}	$\psi_{SN}^{0.4}$	$\psi_{SON}^{0.1}$
Banana	0.130	0.130	0.027	0.027	0.030	0.007	0.012	0.022	0.022	0.022	0.022
Blood	0.445	0.422	0.377	0.377	0.372	0.350	0.372	0.327	0.327	0.340	0.327
Cancer	0.067	0.065	0.077	0.077	0.057	0.060	0.067	0.060	0.057	0.052	0.055
Higle	0.167	0.170	0.050	0.050	0.072	0.047	0.052	0.045	0.047	0.047	0.047
Ion	0.060	0.057	0.062	0.062	0.037	0.110	0.127	0.040	0.037	0.040	0.037
Liver	0.417	0.417	0.412	0.412	0.337	0.315	0.335	0.402	0.325	0.352	0.325
Mam	0.485	0.485	0.377	0.377	0.372	0.350	0.380	0.365	0.357	0.362	0.362
Park	0.052	0.060	0.015	0.015	0.012	0.052	0.057	0.027	0.012	0.015	0.010
Pima	0.425	0.422	0.435	0.435	0.405	0.377	0.390	0.415	0.400	0.407	0.397
Sonar	0.052	0.042	0.047	0.047	0.010	0.070	0.072	0.015	0.012	0.020	0.012
Statlog	0.100	0.105	0.090	0.090	0.067	0.157	0.155	0.092	0.090	0.085	0.090
Mean rank	9.09	8.90	7.09	7.09	4.36	5.00	7.36	4.63	2.27	3.36	2.09

for the case where there is no normalization of the posteriori probability functions. For this case the optimal value of the parameter α is 0.1. Classifiers with the selected values of parameter α were compared with the base classifiers and ensemble methods based on the majority voting and sum methods. The results of classification with the mean ranks obtained by the Friedman test are presented in Table 4. To compare the results the post-hoc Nemenyi test was used [24]. The critical difference for this test at $p = 0.05$ is equal to $CD = 4.51$. Since the difference between Ψ_{Sum} and the proposed algorithms $\Psi_{S_N}^{0.4}, \Psi_{S_{ON}}^{0.1}$ is already smaller than 4.51, we can conclude that the post-hoc test is not powerful enough to detect any significant differences between these algorithms. However, the proposed selection of the posteriori probability functions algorithm achieved the best result, which is the lowest average rank (Table 3).

4 Conclusion

This paper discusses the classifier selection for the binary classification task. The proposal in the paper process concerns the selection of the posteriori probability functions. The paper presents two approaches. In one of them normalization of the posteriori probability functions is carried out. In the second case under consideration, normalization is not performed. The distributed computing approaches enable the efficient and parallel processing of the complicated data analysis task, also in the context of classification systems with multiple classifiers [20]. The methods of classifier selection presented in the paper show the ability to work in a parallel and distributed environment. Parallel processing provides the possibility to speed up the selection of the posteriori probability functions, whose results are needed to make the decision by the classifier ensemble. Additionally, the proposed approach can be applied in various practical tasks involving multiple elementary classification tasks [6, 8, 9]. In the paper several experiments were carried out on the data sets available from the UCI repository and on the synthetical data sets. The aim of the experiments was to compare the proposed selection method algorithms with the base classifiers and ensemble classifiers based on the majority voting and sum methods. For the proposed selection method, we obtained improvement of the classification quality measured by average values from the Friedman test. However, the difference in average ranks is too small to detect statistically significant differences between the proposed selection method and the ensemble method based on the sum rule.

Acknowledgments This work was supported by the Polish National Science Center under the grant no. DEC-2013/09/B/ST6/02264 and by the statutory funds of the Department of Systems and Computer Networks, Wroclaw University of Technology.

References

1. Bishop, C.M.: *Pattern Recognition and Machine Learning* (Information Science and Statistics). Springer, New York (2006)
2. Burduk, R.: Classifier fusion with interval-valued weights. *Pattern Recognit. Lett.* **34**(14), 1623–1629 (2013)
3. Britto, A.S., Sabourin, R., Oliveira, L.E.S.: Dynamic selection of classifiers—a comprehensive review. *Pattern Recognit.* **47**(11), 3665–3680 (2014)
4. Cyganek, B.: One-class support vector ensembles for image segmentation and classification. *J. Math. Imaging Vis.* **42**(2–3), 103–117 (2012)
5. Cavalin, P.R., Sabourin, R., Suen, C.Y.: Dynamic selection approaches for multiple classifier systems. *Neural Comput. Appl.* **22**(3–4), 673–688 (2013)
6. Cyganek, B., Woźniak, M.: *Vehicle Logo Recognition with an Ensemble of Classifiers*. Lecture Notes in Computer Science, vol. 8398, pp. 117–126. Springer, Berlin (2014)
7. Didaci, L., Giacinto, G., Roli, F., Marcialis, G.L.: A study on the performances of dynamic classifier selection based on local accuracy estimation. *Pattern Recognit.* **38**, 2188–2191 (2005)
8. Frejlichowski, D.: *An Algorithm for the Automatic Analysis of Characters Located on Car License Plates*. Lecture Notes in Computer Science, vol. 7950, pp. 774–781. Springer, Berlin (2013)
9. Forczmański, P., Łabędź, P.: *Recognition of Occluded Faces Based on Multi-subspace Classification*. Lecture Notes in Computer Science, vol. 8104, pp. 148–157. Springer, Berlin (2013)
10. Frank, A., Asuncion, A.: *UCI Machine Learning Repository* (2010)
11. Giacinto, G., Roli, F.: An approach to the automatic design of multiple classifier systems. *Pattern Recognit. Lett.* **22**, 25–33 (2001)
12. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003)
13. Ho, T.K., Hull, J.J., Srihari, S.N.: Decision combination in multiple classifier systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(1), 66–75 (1994)
14. Jackowski, K., Woźniak, M.: Method of classifier selection using the genetic approach. *Expert Syst.* **27**(2), 114–128 (2010)
15. Jackowski, K., Krawczyk, B., Woźniak, M.: Improved adaptive splitting and selection: the hybrid training method of a classifier based on a feature space partitioning. *Int. J. Neural Syst.* **24**(3) (2014)
16. Kuncheva, L.I.: A theoretical study on six classifier fusion strategies. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(2), 281–286 (2002)
17. Kuncheva, L.I.: *Combining Pattern Classifiers: Methods and Algorithms*. Wiley, Hoboken (2004)
18. Kittler, J., Alkoot, F.M.: Sum versus vote fusion in multiple classifier systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(1), 110–115 (2003)
19. Lam, L., Suen, C.Y.: Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Trans. Syst., Man, Cybern., Part A* **27**(5), 553–568 (1997)
20. Przewoźniczek, M., Walkowiak, K., Woźniak, M.: Optimizing distributed computing systems for k-nearest neighbours classifiers—evolutionary approach. *Logic J. IGPL* **19**(2), 357–372 (2010)
21. Rejer, I.: Genetic Algorithms in EEG Feature Selection for the Classification of Movements of the Left and Right Hand. In: *Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013. Advances in Intelligent Systems and Computing*, vol. 226, pp. 581–590. Springer, Heidelberg (2013)
22. Ranawana, R., Palade, V.: Multi-classifier systems: review and a roadmap for developers. *Int. J. Hybrid Intell. Syst.* **3**(1), 35–61 (2006)
23. Ruta, D., Gabrys, B.: Classifier selection for majority voting. *Inf. Fusion* **6**(1), 63–81 (2005)
24. Smętek, M., Trawiński, B.: Selection of heterogeneous fuzzy model ensembles using self-adaptive genetic algorithms. *New Gener. Comput.* **29**(3), 309–327 (2011)

25. Suen, C.Y., Legault, R., Nadal, C.P., Cheriet, M., Lam, L.: Building a new generation of handwriting recognition systems. *Pattern Recognit. Lett.* **14**(4), 303–315 (1993)
26. Trawiński, K., Cordon, O., Quirin, A.: A study on the use of multiobjective genetic algorithms for classifier selection in furia-based fuzzy multiclassifiers. *Int. J. Comput. Intell. Syst.* **5**(2), 231–253 (2012)
27. Ulas, A., Semerci, M., Yildiz, O.T., Alpaydin, E.: Incremental construction of classifier and discriminant ensembles. *Inf. Sci.* **179**(9), 1298–1318 (2009). Apr
28. Wołoszyński, T., Kurzyński, M.: A probabilistic model of classifier competence for dynamic ensemble selection. *Pattern Recognit.* **44**(10–11), 2656–2668 (2011)