classifier fusion, dynamic ensemble selection, multiple classifier system

Paulina BACZYŃSKA[1], Robert BURDUK[1]

DYNAMICAL ENSEMBLE SELECTION - EXPERIMENTAL ANALYSIS ON HOMOGENOUS POOL OF CLASSIFIERS

The paper presents the dynamic ensemble selection based on the analysis of the decision profiles. These profiles are obtained from a posteriori probability functions returned from the base classifiers during the training process. Presented in the paper dynamic ensemble selection algorithms are dedicated to the binary classification task. In order to verify these algorithms, a number of experiments have been carried out on several medical data sets. The proposed dynamic ensemble selection is experimentally compared against the ensemble with the sum fusion method. As base classifiers we used the pool of homogeneous classifiers. The obtained results are promising because we could improve the classification accuracy of the ensemble classifier.

1. INTRODUCTION

Classification is one of the important steps in pattern recognition, which belongs to machine learning fields [1]. The classification task can be accomplished by a single classifier or by a team of classifiers. In the literature, the use of the multiple classifiers for a decision problem is known as the multiple classifier systems (MCS) or the ensemble of classifiers EoC [4], [8]. The construction of MSC consists of three phases: generation, selection and integration [2]. In the second phase, which is discussed in this paper, one or a subset of the base classifiers is selected to make the final decision which it is to assign an object to the class label.

The output of an individual classifier can be divided into three types [14].

- The abstract level – the classifier ψ assigns the unique label j to a given input x.
- The rank level – in this case for each input (object) 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 the measurement value that addresses the degree of assigning the class label to the given input x. An example of such a representation of the output is a posteriori probability returned by Bayes classifier.

According to these three types of outputs of the base classifier, various problems of the combination function of classifier' outputs are considered. The problems studied in [15], [19] belong to the abstract level. The combining outputs for the rank level are presented in [9] and problems studied in [12], [13] belong to the last level.

[1] Department of Systems and Computer Networks, Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland e-mail: robert.burduk@pwr.edu.pl

The selection of classifiers is one of the important problems in the creation of EoC [10], [18]. This task is related to the choice of a set of classifiers from all the available pool of classifiers. Here you can distinguish between the static or dynamic selection [16]. In the static classifier selection one set of classifiers is selected to create an 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 [17]. In the dynamic classifier selection for each unknown sample a specific subset of classifiers is selected [3]. It means that we are selecting different EoCs for different objects from the testing set. In this type of the classifier selection, the classifier is chosen and assigned to the sample based on different features [20] or different decision regions [6], [11].

In this work we present the dynamic selection of a posteriori probability functions (PPFs). In particular we make the experimental analysis of the homogenous pool of classifiers based on the several medical data sets.

The text is organized as follows: after this introduction, in Section II the idea of EoC is presented. Section III contains the description of the proposed dynamic selection of PPFs. The experimental results on medical data sets are presented in Section IV.

2. ENSEMBLE OF CLASSIFIERS

Let us assume that we possess K of different classifiers $\Psi_1, \Psi_2, \ldots, \Psi_K$. Each classifier transform a feature vector x to a class label i, i.e. $\Psi(x) \to i$. Such a set of classifiers, which is constructed on the basis of the same learning sample is called an ensemble of classifiers or a combined classifier. However, any of Ψ_i classifiers is described as a component or base classifier. As a rule K is assumed to be an odd number and each of Ψ_i classifiers makes an independent decision. As a result, of all the classifiers' action, their K responses are obtained. Having at the disposal a set of base classifiers one should determine the procedure of making the ultimate decision regarding the allocation of the object to one of the available classes. It implies that the output information from all K component classifiers is applied to make the ultimate decision.

2.1. COMBINATION FUNCTION OF CLASSIFIERS OUTPUTS

In this work we consider the situation when each base classifier returns the estimation of a posteriori probability. This means that outputs of all the base classifiers are at the measurement level [14]. Let us denote a posteriori probability estimation by $\hat{p}_k(i|x)$, $k = 1, 2, \ldots, K$, $i = 1, 2, \ldots, M$, where M is the number of the class labels. One of the possible possible approaches consists in linear combination of such outputs. This method makes use of the linear function such as Sum, Prod or Mean for the combination of the outputs. In the sum method the score of the group of classifiers is based on the application of the following sums:

$$s_i(x) = \sum_{k=1}^{K} \hat{p}_k(i|x), \qquad i = 1, 2, \ldots, M. \tag{1}$$

The final decision of the group of classifiers is made following the maximum rule:

$$\Psi_S(x) = \arg\max_i s_i(x), \qquad i = 1, 2, \ldots, M. \tag{2}$$

Similarly, in the mean method we use the following formula:

$$m_i(x) = \frac{1}{K}\sum_{k=1}^{K} \hat{p}_k(i|x), \qquad i = 1, 2, \ldots, M, \tag{3}$$

and in the product method the following relation is used:

$$p_i(x) = \prod_{k=1}^{K} \hat{p}_k(i|x), \qquad i = 1, 2, \ldots, M. \tag{4}$$

Now the final decision of the ensemble of classifiers is made according to the mean rule:

$$\Psi_M(x) = \arg\max_i m_i(x), \qquad i = 1, 2, \ldots, M, \tag{5}$$

or the product rule:

$$\Psi_P(x) = \arg\max_i p_i(x), \qquad i = 1, 2, \ldots, M. \tag{6}$$

In the presented methods (2), (5), (6) discrimination functions obtained from the individual classifiers take an equal part in building the combined classifier.

2.2. DYNAMIC ENSEMBLE SELECTION

Being given K classifiers from the initial pool of classifiers we select a posteriori probability functions (PPFs) returned by this pool. The selected PPFs are integrated and are used to built the ensemble. The final decision is made on the basis of the dynamically selected PPFs. It means that the selection is performed for each new object (from the testing sets).

Now we present the algorithm for the selection of PPFs. For the binary classification task and for K base classifier their outputs are arranged in the decision profile:

$$DP(x) = \begin{bmatrix} \hat{p}_1(1|x) & \hat{p}_1(2|x) \\ \vdots & \vdots \\ \hat{p}_K(1|x) & \hat{p}_K(2|x) \end{bmatrix}. \tag{7}$$

During learning of the base classifiers we obtain m decision profiles, where m is the number of objects from the learning set. In the first stage of the proposed selection of PPFs algorithm we calculate the decision scheme according to the formula:

$$DS = \begin{bmatrix} \widehat{ds}_{11} & \widehat{ds}_{12} \\ \vdots & \vdots \\ \widehat{ds}_{K1} & \widehat{ds}_{K2} \end{bmatrix}, \tag{8}$$

where

$$\widehat{ds}_{ki} = \widehat{\widehat{ds}}_{ki} + \beta\sqrt{\frac{\sum_{n=1}^{m}(I(\Psi_k(x_n) = i_n)\,\hat{p}_k(i_n|x_n) - \widehat{\widehat{ds}}_{ki})^2}{m-1}} \tag{9}$$

and

$$\widehat{\widehat{ds}}_{ki} = \frac{\sum_{n=1}^{m} I(\Psi_k(x_n) = i_n)\,\hat{p}_k(i_n|x_n)}{\sum_{n=1}^{m} I(\Psi_k(x_n) = i_n)}. \tag{10}$$

The $I(\cdot)$ is the indicator function, which means that its value is equal to one in the case of the correct classification of the object x_n by Ψ_k algorithm. This ensures that $\widehat{ds}_{ki}$ is calculated only from those PPFs for which the classifier k did not make an error.

The above decision scheme is used in the selection of PPFs from the decision profile for the new object according to the formula:

$$\text{if } \hat{p}_k(i|x) < \widehat{ds}_{ki} \text{ then } \hat{p}_k(i|x) = null, k = 1, ..., K, \ i = 1, 2. \tag{11}$$

The obtained decision profile, designated as DP_{DS}, for the new object contains the selected PPFs. Based on the DP_{DS} we can use the various algorithms for the integration of PPFs. In experimental studies we use the sum method (2) to make the final decision by the ensemble classifier after the selection of PPFs. The algorithm using this method is denoted as Ψ_{DS}.

In the second version of the proposed dynamic ensemble selection algorithm the normalization is carried out. The normalization is performed for each label class i according to the rule:

$$\widehat{p'}_k(i|x) = \frac{\hat{p}_k(i|x) - \min(\hat{p}_1(i|x), ..., \hat{p}_k(i|x))}{\max(\hat{p}_1(i|x), ..., \hat{p}_k(i|x)) - \min(\hat{p}_1(i|x), ..., \hat{p}_k(i|x))}, \qquad k \in K. \tag{12}$$

The other steps of the algorithm in the training process are the same as described above. In the testing process the normalization is carried out for each decision profile of the testing object similarly to the formula (12). The algorithm using method with normalization is denoted as Ψ_{DS-N}.

3. EXPERIMENTAL STUDIES

In the experiential research 6 medical data sets were tested. The data sets come from UCI repository [7]. The numbers of attributes and available examples of the investigated data sets are introduced in Tab. 1. A set of all the available features was used for all data sets. In the experiment we use the binary data sets with class labels 0 and 1. In Tab. 1 is presented the ratio $(0/1)$.

The aim of the experiments was to compare the quality of classifications of the proposed dynamic selection algorithms with the ensemble classifier which uses the sum method. The ensemble classifier in our research is composed of homogeneous classifiers. In the experiments we use the ensemble classifier, which consists of different $k-NN$ or SVM classifiers.

Table 1. Description of data sets selected for the experiments.

Data set	example	attribute	ration $(0/1)$
Blood	748	5	3.2
Breast Cancer Wisconsin	699	10	1.9
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

The research assumes that the ensemble of classifiers is composed of 5 elementary classifiers. Tab. 2 presents the obtained results for the case when we use $k-NN$ as base classifiers. Tab. 3 show the results of classification for case when we use the SVM as base classifiers. Tab. 2- 3 show classification error and the average ranks obtained by the Friedman test [5]. The results are obtained via 10-fold-cross-validation method.

In the algorithms comparison we use the average ranks. The obtained results are promising because we could improve the classification accuracy of homogenous pool of the base classifiers. In our experiments it can be seen for the algorithm labeled as $\Psi_{DS-N}^{\beta=-1}$. It has always better classification accuracy compared to the ensemble algorithm based on the sum rule Ψ_S. The algorithm $\Psi_{DS-N}^{\beta=-1}$ has a lower value of the average rank than the algorithm Ψ_S.

Table 2. Classification error for $k - NN$ base classifiers.

Data set	Ψ_S	$\Psi_{DS-N}^{\beta=-1}$	$\Psi_{DS-N}^{\beta=0}$	$\Psi_{DS}^{\beta=-1}$	$\Psi_{DS}^{\beta=0}$
Blood	0.282	0.267	0.223	0.279	0.323
Cancer	0.046	0.021	0.020	0.047	0.060
Liver	0.334	0.343	0.344	0.360	0.334
Mammographic	0.237	0.231	0.183	0.249	0.259
Parkinson	0.200	0.236	0.157	0.211	0.200
Pima	0.281	0.279	0.240	0.279	0.305
Ave. rank	3.3	1	2.7	4.5	3.5

Table 3. Classification error for SVM base classifiers.

Data set	Ψ_S	$\Psi_{DS-N}^{\beta=-1}$	$\Psi_{DS-N}^{\beta=0}$	$\Psi_{DS}^{\beta=-1}$	$\Psi_{DS}^{\beta=0}$
Blood	0.214	0.193	0.169	0.210	0.291
Cancer	0.030	0.053	0.046	0.030	0.032
Liver	0.270	0.179	0.087	0.274	0.387
Mammographic	0.216	0.197	0.182	0.224	0.220
Parkinson	0.216	0.094	0.079	0.200	0.205
Pima	0.259	0.240	0.267	0.264	0.284
Ave. rank	2.9	1.8	2.3	4.3	3.6

Additionally, the obtained results show the the proposed algorithm with normalization is always better than the proposed algorithm without normalization. In future work the proposed dynamic ensemble selection method should be examined on a larger set of base classifiers as well as on homogeneous pool of classifiers.

4. CONCLUSION

This paper presents the dynamic ensemble classifier selection algorithm dedicated to the binary classification task. The presented dynamic selection algorithm is based on the analysis of the decision profiles returned from the base classifiers during the training process. Experimental studies were carried out on the several medical data sets available from the UCI repository. They show that using the proposed in the work dynamic ensemble selection method can improve the quality of classification in the cases of the homogeneous ensemble classifiers. In the future work the heterogeneous pool of classifiers should be tested.

ACKNOWLEDGEMENT

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.

BIBLIOGRAPHY

[1] BISHOP C. Bishop pattern recognition and machine learning. 2006. Springer, New York.

[2] BRITTO A. S., SABOURIN R., OLIVEIRA L. E. Dynamic selection of classifiersa comprehensive review. Pattern Recognition, 2014, Vol. 47. Elsevier, pp. 3665–3680.

[3] CAVALIN P. R., SABOURIN R., SUEN C. Y. Dynamic selection approaches for multiple classifier systems. Neural Computing and Applications, 2013, Vol. 22. Springer, pp. 673–688.

[4] CYGANEK B. One-class support vector ensembles for image segmentation and classification. Journal of Mathematical Imaging and Vision, 2012, Vol. 42. Springer, pp. 103–117.

[5] DEMŠAR J. Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research, 2006, Vol. 7. JMLR. org, pp. 1–30.

[6] DIDACI L., GIACINTO G., ROLI F., MARCIALIS G. L. A study on the performances of dynamic classifier selection based on local accuracy estimation. Pattern Recognition, 2005, Vol. 38. Elsevier, pp. 2188–2191.

[7] FRANK A., ASUNCION A., ET AL. Uci machine learning repository. 2010.
[8] GIACINTO G., ROLI F. An approach to the automatic design of multiple classifier systems. Pattern recognition letters, 2001, Vol. 22. Elsevier, pp. 25–33.
[9] HO T. K., HULL J. J., SRIHARI S. N. Decision combination in multiple classifier systems. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1994, Vol. 16. IEEE, pp. 66–75.
[10] JACKOWSKI K., KRAWCZYK B., WOŹNIAK M. Improved adaptive splitting and selection: The hybrid training method of a classifier based on a feature space partitioning. International journal of neural systems, 2014, Vol. 24. World Scientific, p. 1430007.
[11] JACKOWSKI K., WOZNIAK M. Method of classifier selection using the genetic approach. Expert Systems, 2010, Vol. 27. Wiley Online Library, pp. 114–128.
[12] KITTLER J., ALKOOT F. M. Sum versus vote fusion in multiple classifier systems. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2003, Vol. 25. IEEE, pp. 110–115.
[13] KUNCHEVA L. I. A theoretical study on six classifier fusion strategies. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, no. 2. IEEE, pp. 281–286.
[14] KUNCHEVA L. I. Combining pattern classifiers: methods and algorithms. 2004. John Wiley & Sons.
[15] LAM L., SUEN C. Y. Application of majority voting to pattern recognition: an analysis of its behavior and performance. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 1997, Vol. 27. IEEE, pp. 553–568.
[16] RANAWANA R., PALADE V. Multi-classifier systems: Review and a roadmap for developers. Int. J. Hybrid Intell. Syst., 2006, Vol. 3. pp. 35–61.
[17] RUTA D., GABRYS B. Classifier selection for majority voting. Information fusion, 2005, Vol. 6. Elsevier, pp. 63–81.
[18] SMETEK M., TRAWIŃSKI B. Selection of heterogeneous fuzzy model ensembles using self-adaptive genetic algorithms. New Generation Computing, 2011, Vol. 29. Springer, pp. 309–327.
[19] SUEN C. Y., LEGAULT R., NADAL C., CHERIET M., LAM L. Building a new generation of handwriting recognition systems. Pattern Recognition Letters, 1993, Vol. 14. Elsevier, pp. 303–315.
[20] WOLOSZYNSKI T., KURZYNSKI M. A probabilistic model of classifier competence for dynamic ensemble selection. Pattern Recognition, 2011, Vol. 44. Elsevier, pp. 2656–2668.