Classifier fusion, static classifiers selection, multiple classifier system

Robert BURDUK[1]

THE PROPOSAL OF CALCULATION CLASSIFIER WEIGHTS FOR AN ASSEMBLY OF CLASSIFIERS

The selection of classifiers is one of the important problems in the creation of ensemble of classifiers. The paper presents the static selection in which a new method of calculating the weights of individual classifiers is used. The obtained weights can be interpreted in the context of the interval logic. It means that the particular weights will not be provided precisely but their lower and upper values will be used. A number of experiments have been carried out on several medical data sets.

1. INTRODUCTION

The selection of classifiers is one of the important problems in the creation of ensemble of classifiers (EoC) [6], [11]. This task is related to the choice of a set of classifiers of all the available pool of classifiers. In 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 [10]. In the dynamic classifier selection for each unknown sample a specific subset of classifiers is selected [2]. It means that we are selecting different EoCs for different object from the testing set. In this type of the classifier selection, the classifier is chosen and assigned to the sample based on different features [12] or different decision regions [3], [7].

In this work we will consider static approach to built the EoC. In detail we propose the new method to select the classifiers from the available pool. This method is based on the correction of base classifiers and can be interpreted in the contents of the interval logic. The presented results are compared with the oracle concept [4] and base classifiers. The oracle classifier is used as the possible upper limit of classification accuracy of the EoC. As a fusion function for classifier outputs we use the sum and weighted sum methods. The advantage of EoC proposed in the paper is possibility to work in parallel and distributed environment. Only the process of calculating the weights of classifiers requires data from the outputs of all other base classifiers.

The text is organized as follows: in Section II the ensemble of classifiers and combination functions of classifiers outputs are presented. Section III contains the new method for assigning

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

weights of individual base classifiers. Section IV includes the description of research experiments comparing the suggested algorithms with base classifiers, oracle, sum, product and mean methods. Finally, conclusions from the experiments are presented.

2. ENSEMBLE OF CLASSIFIERS

Let us assume that we possess K of different classifiers $\Psi_1, \Psi_2, \ldots, \Psi_K$. 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 [8]. 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 linear function like 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. Also, the weighted versions of these methods can be created. In this approach each of the classifiers has an allocated weight, which is taken into account whiles reaching the final decision of the group of classifiers. Weights depend largely on the quality of their base classifiers. In the case when each classifier has one weight for all the possible classes an adequate group classification formula for the sum method is presented as follows:

$$sw_i(x) = \sum_{k=1}^{K} w_k * \hat{p}_k(i|x), \qquad i = 1, 2, \ldots, M, \tag{7}$$

where $w_k = 1 - Pe_{\Psi_k}$, and Pe_{Ψ_k} is the empirical error of Ψ_k classifier estimated on the testing set. In the case when the error is estimated on the learning set, we can talk about the estimation error based on the resubstitution method. Then w_k weight of each component classifier is calculated depending on the:

$$w_k = \frac{\sum_{n=1}^{N} I(\Psi_k(x_n) = i, j_n = i)}{N}, \qquad i = 1, 2, \ldots, M, \tag{8}$$

where $I()$ is an indicator function. The N value refers to the number of the learning set observations, which is used for estimating classifiers' weights, and j_n is the class number of the object with n index.

The obtained weights are normalised according to the formula:

$$\sum_{k=1}^{K} w_k = 1, \tag{9}$$

which means that the sum of weights of all classifiers from the ensemble is equal to unity. In this case the final decision of the ensemble of classifiers is the following:

$$\Psi_{wS}(x) = \arg\max_i sw_i(x). \tag{10}$$

For the formula (1) we have $w_k = 1$ for all $k = 1, \ldots, K$. Having defined the weights (8) you can easily use them for other rules (5) and (6).

Another approach to obtain weights, is the calculation them for each class separately. Then the corresponding weight is calculated from the equation:

$$w_{ki} = \frac{\sum_{n=1}^{N} I(\Psi_k(x_n) = i, j_n = i)}{\sum_{n=1}^{N} I(j_n = i)}. \tag{11}$$

In this case in the formula (8) is used w_{ki} instead of w_k. The sum method assumes designation of $\Psi_{wcS}(x)$.

3. STATIC CLASSIFIER SELECTION

We will suggest now the method for determining weights for the individual base classifiers. The values of these weights are the basis for the selection of classifiers. These weights can be seen in the context of the interval logic. It means that the particular weights will not be provided precisely but their lower and upper values will be used. Therefore, each w_k weight of K-component classifier will be represented by the upper $\overline{w}_k^{ss}$ and lower $\underline{w}_k^{ss}$ value.

Let us now present the method for calculating weights for individual classifiers in which the basis for the determination of upper and lower values will be the correct classification of component classifiers. Having at the disposal a group of K component classifiers $\Psi_1, \Psi_2, \ldots, \Psi_K$ we ascribe at the learning set the probability of the correct classification Pc_{Ψ_k} for each of them. The upper value $\overline{w}_k^{ss}$ of k-classifier weight refers to the situation in which k-classifier was correct, while the other committee classifiers proved the correct prediction. The lower value $\underline{w}_k^{ss}$ describes the situation in which k-classifier made errors, while the other committee classifiers did not make any errors. The upper value is obtained from the dependence:

$$\overline{w}_k^{ss} = \frac{\sum_{n=1}^{N} UC_k^{ss_n}}{\arg\max_{l \in \mathcal{K}} \sum_{n=1}^{N} UC_l^{ss_n}}, \tag{12}$$

where

$$UC_k^{ss_n} = \frac{I(\Psi_k(x_n) = j_n)}{\sum_{l=1, l \neq k}^{K} I(\Psi_l(x_n) = j_n)}. \tag{13}$$

However, the lower value is obtained from the dependence:

$$\underline{w}_k^{ss} = \frac{\sum_{n=1}^{N} LC_k^{ss_n}}{\arg\max_{l \in \mathcal{K}} \sum_{n=1}^{N} LC_l^{ss_n}}, \tag{14}$$

where

$$LC_k^{ss_n} = \frac{I(\Psi_k(x_n) \neq j_n)}{\sum_{l=1, l \neq k}^{K} I(\Psi_l(x_n) = j_n)}. \tag{15}$$

Similarly, as in equation (11), we can calculated for each class the appropriate lower and upper values of weights.

3.1. CLASSIFIER SELECTION

Given K classifiers from the initial pool of classifiers now we select L, $L \leq K$ classifiers to the ensemble. The final decision is made on the basis of L classifiers. In the selection process, we set the value of L. Then choose from the available pool L classifiers with the largest coefficients w_k. In the set of all $w_k, k = 1, \ldots, K$ we find the w_k^L with the L-th largest value. Now we define the set $w^{SS} = \{w_k : w_k \geq w_k^L\}$. It means that we select L best classifiers. The advantage of this method is that it is very cheap computationally [10].

First we create coefficient α_k^{SS} according to the formula:

$$\alpha_k^{SS} = \begin{cases} w_k & \text{if } w_k \in w^{SS} \\ 0 & \text{otherwise} \end{cases}. \tag{16}$$

If we use the sum method for the final combination of classifier outputs, then the score of the selected group of classifiers is the following:

$$s_i^{SS}(x) = \sum_{k=1}^{K} \alpha_k^{SS} * \hat{p}_k(i|x), \qquad i = 1, 2, \ldots, M. \tag{17}$$

The final decision of the selected group of classifiers is made according to the formula:

$$\Psi_{wS}^{SS}(x) = \arg\max_{i} s_i^{SS}(x). \tag{18}$$

Before making the final decision the coefficients $\alpha_1^{SS}, ..., \alpha_K^{SS}$ are normalised to unity.

4. EXPERIMENTAL STUDIES

In the experiential research 4 medical data sets were tested. The first set refers to the acute abdominal pain diagnosis problem and comes from the Surgical Clinic Wroclaw Medical Academy. The other three data sets come from UCI repository [5]. A set of all the available features was used for all data sets, however, for the acute abdominal pain data set the selection of features has been made in accordance with the suggestions from another work on the topic [1], [9].

The numbers of attributes, classes and available examples of the investigated data sets are introduced in Tab. 1. The aim of the experiments was to compare the quality of classifications of the proposed static selection method algorithms with the oracle method, sum method and base classifiers. The oracle strategy correctly classifies the test sample if any of the classifier from the ensemble predicts the correct label for this sample.

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

Data set	example	attribute	class
Acute Abdominal Pain	476	31	8
Breast Tissue	106	10	6
Dermatology	366	33	6
Pima Indians Diabetes	768	8	2

The research assumes that the group of classifiers is composed of 7 elementary classifiers. Three of them work according to the $k - NN$ rule where the k parameter is from the set $k \in 3, 5, 7$. For the four remaining base classifiers the decision trees are used, with the number

Table 2. Classification error for base classifiers and oracle classifier.

Data set	Ψ_1	Ψ_2	Ψ_3	Ψ_4	Ψ_5	Ψ_6	Ψ_7	oracle
Acute	0.161	0.166	0.166	0.171	0.161	0.159	0.164	0.042
Breast	0.43	0.459	0.5	0.319	0.373	0.305	0.405	0.103
Dermat.	0.125	0.137	0.152	0.078	0.109	0.07	0.068	0.003
Pima	0.261	0.254	0.241	0.269	0.256	0.265	0.264	0.078

Table 3. Classification error for classifiers with weights calculated as the fraction of correctly classified objects - upper values of weights.

Data set	L	Ψ_S	Ψ_P	Ψ_M	Ψ_{wS}^{SS}	Ψ_{wP}^{SS}	Ψ_{wM}^{SS}	Ψ_{wcS}^{SS}	Ψ_{wcP}^{SS}	Ψ_{wcM}^{SS}
	7	0.160	0.159	0.159	0.158	0.161	0.159	0.161	0.168	0.157
Acute	5				0.158	0.168	0.149	0.158	0.158	0.164
	3				0.161	0.154	0.158	0.148	0.148	0.158
	7	0.319	0.37	0.319	0.311	0.37	0.311	0.319	0.37	0.319
Breast	5				0.295	0.378	0.295	0.311	0.368	0.316
	3				0.276	0.319	0.276	0.308	0.368	0.316
	7	0.051	0.08	0.051	0.051	0.08	0.051	0.051	0.08	0.051
Dermat.	5				0.055	0.08	0.055	0.053	0.077	0.056
	3				0.066	0.07	0.066	0.054	0.072	0.059
	7	0.235	0.238	0.235	0.233	0.238	0.233	0.233	0.237	0.233
Pima	5				0.233	0.231	0.233	0.229	0.231	0.229
	3				0.238	0.235	0.238	0.233	0.238	0.233

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 are used. The results are obtained via 10-fold-cross-validation method. Tab. 2 presents the obtained results for seven base classifiers and for oracle classifier. Tab. 3 show the results of classification for the initial ensemble of classifiers, when $L = 7$ and results after the classifier selection process, when $L = 5$ or $L = 3$.

The obtained results confirm the validity of the application for the proposed method for classifiers selection.

5. CONCLUSION

This paper presents static classifier selection algorithm in which a new method of calculating the weights is used. Experimental studies were carried out on the medical data sets available from the UCI repository and one data sets form the Surgical Clinic Wroclaw Medical Academy. They show that using the proposed in the work method for calculating weights can improvement the classification quality measured by average error.

In the future work the additional sets of data should be used to perform statistical analysis of the results. Additional tests may also include a linear combination of the proposed in this article upper and lower values of weights.

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] BURDUK R., WOŹNIAK M., Different decision tree induction strategies for a medical decision problem. Central European Journal of Medicine, 2012, Vol. 7 (2), pp. 183–193.

[2] CAVALIN P. R., SABOURIN R., SUEN. C. Y., Dynamic selection approaches for multiple classifier systems. Neural Computing and Applications, 2013, Vol. 22 (3-4), pp. 673–688.

[3] 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, 11/2005, Vol. 38, pp. 2188–2191.

[4] DOS SANTOS E. M., SABOURIN R., Classifier ensembles optimization guided by population oracle. In IEEE Congress on Evolutionary Computation, 2011, pp. 693–698.

[5] FRANK A., ASUNCION A., UCI machine learning repository, 2010.

[6] 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 (03).

[7] JACKOWSKI K., WOŹNIAK M., Method of classifier selection using the genetic approach. Expert Systems, 2010, Vol. 27 (2), pp. 114–128.

[8] KUNCHEVA L. I., Combining Pattern Classifiers: Methods and Algorithms. John Wiley and Sons, Inc., 2004.

[9] KURZYŃSKI M. W., Diagnosis of acute abdominal pain using a three-stage classifier. Computers in biology and medicine, 1987, Vol. 17 (1), pp. 19–27.

[10] RUTA D., GABRYS B., Classifier selection for majority voting. Information Fusion, 2005, Vol. 6 (1), pp. 63–81.

[11] SMETEK M., TRAWIŃSKI B., Selection of heterogeneous fuzzy model ensembles using self-adaptive genetic algorithms. New Generation Comput., 2011, Vol. 29 (3), pp. 309–327.

[12] WOLOSZYNSKI T, KURZYŃSKI M. W., A probabilistic model of classifier competence for dynamic ensemble selection. Pattern Recognition, 2011, Vol. 44 (10-11), pp. 2656–2668.